华章数学译丛

22

HZ BOOKS

PEARSON Prentice Hall

A First Course in Probability

(Sixth Edition)

概率论基础教程

(原书第6版)

(美) Sheldon Ross 南加州大学 著

赵选民 等译



机械工业出版社
China Machine Press

(原书第6版)

概率论基础教程

本书是一本概率论的入门教材，系统介绍了概率论的基础理论及应用，在取材、结构和写作方法等方面具有鲜明的特点。通过例题阐述概率论的基本概念与方法是本书的一大特色。作者独具匠心地选择和编排了大量例题与习题，这些内容约占全书的三分之二。通过这些例题和习题，读者可以了解概率论在各个领域的广泛应用，如基因、彩票、法庭判决、NBA选秀等。

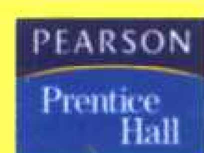
作 者 简 介



Sheldon Ross 于1968年在斯坦福大学获得统计学博士学位，现为南加州大学工业工程与系统工程系教授，曾执教于加州大学伯克利分校工业工程与运筹学系。除本书外，他还著有《数理金融初步》（该书中文版、影印版已由机械工业出版社引进出版）、《Simulation》等书。另外，他还发表了大量有关概率与统计方面的学术论文，创办了《Probability in the Engineering and Informational Sciences》杂志并一直担任主编。他是数理统计学会会员，荣获过美国科学家Humboldt奖。

A First Course in Probability

(Sixth Edition)



www.PearsonEd.com





华章网站 http://www.hzbook.com

网上购书：www.china-pub.com

投稿热线：(010) 88379604

购书热线：(010) 68995259, 68995264

读者信箱：hzjsj@hzbook.com





华章图书

ISBN 7-111-18378-9/O・477

定价：42.00 元

封面设计・杨宇梅

华章数学译丛

22

A First Course in Probability

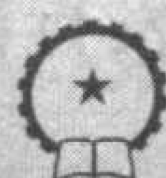
(Sixth Edition)

概率论基础教程

（原书第6版）

(美) Sheldon Ross 著
南加州大学

赵选民 等译



机械工业出版社
China Machine Press

本书系统介绍了概率论的基础理论及应用，主要内容包括组合分析、概率论的公理、条件概率与独立性、随机变量及其分布、数学期望、极限定理、随机模拟等. 另外，作者精心选择了大量的例题和习题，揭示了概率论在各个领域的广泛应用.

本书通俗易懂，可作为高等院校相关专业概率论课程的教材或教学参考书.

本书版权登记号：图字：01-2004-2894

图书在版编目（CIP）数据

概率论基础教程(原书第6版)/(美)罗斯(Ross,S.)著；赵选民等译. —北京：机械工业出版社，2006.4

(华章数学译丛)

书名原文：A First Course in Probability, Sixth Edition

ISBN 7-111-18378-9

Ⅰ. 概…　Ⅱ. ①罗…　②赵…　Ⅲ. 概率论－教材　Ⅳ. O21

中国版本图书馆 CIP 数据核字(2006)第004680号

机械工业出版社(北京市西城区百万庄大街22号　邮政编码　100037)

责任编辑：迟振春

北京牛山世兴印刷厂印刷 · 新华书店北京发行所发行

2006年4月第1版第1次印刷

787mm×1020mm 1/16 · 23.25 印张

定价：42.00元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换

本社购书热线：(010)68326294

译 者 序

概率论是研究随机现象规律的数学分支，始于20世纪30年代，其发展源于它自身逻辑基础的建立和科学技术、社会实践的许多实际需要．现在概率论不仅在随机过程、随机分析和极限理论等领域受到广泛关注，而且数理统计、数理金融和生物数学等学科也都密切地与概率论的发展相联系．

本书是一本概率论的入门教材，在取材、结构和写作方法等方面具有鲜明的特点．通过例题阐述概率论的基本概念与方法是本书的一大特色，作者独具匠心地选择和编排了大量例题与习题，这些内容约占全书的三分之二．通过这些例题和习题，读者可以了解概率论在各个领域的广泛应用．

本书通俗易懂，但又不失其科学性、严密性与准确性．对于初学者来说，既能较准确地掌握概率论的基本概念与方法，又不致苦于较深的数学推导所带来的困难与乏味．正如作者在前言中所说．本书可作为高等院校数学、统计学、工程技术以及其他科学（包括计算机科学、社会科学和管理科学）的学生概率论课程的教材或教学参考书，也可供具备初等微积分知识的读者自学参考．

在翻译过程中，我们在忠实原文的基础上，努力使之便于我国读者理解，另外，我们部分地参考了李漳南和杨振明教授于1980年翻译的本书前版，当时的书名为《概率论初级教程》（人民教育出版社），在此谨表谢意．

本书由赵选民主译，另外张未未、李艳玲、苗宇涛、李娟、崔艳丽、解俊山也参加了部分翻译与录入工作．由于译者水平有限，难免存在不妥之处，敬请广大读者不吝赐教．

译　者

前　言

著名的法国数学家和天文学家（曾被称为“法国的牛顿”）皮埃尔·西蒙·拉普拉斯侯爵曾说过：“我们发现，概率论实质上就是被归纳为计算问题的常识，它使我们能正确地评价凭某种直觉所感受到的、往往又不能解释清楚的见解的合理性……值得注意的是，概率论这门起源于机会游戏的科学，终将成为人类知识中最重要的组成部分……对于大多数人来说，生活中最重要的问题正是概率问题．”尽管许多人可能会认为，这位曾对概率论的发展做出巨大贡献的侯爵的话有点言过其实，但是，概率论确实已经成为几乎所有的科学家、工程师、医生、律师和实业家手中的一个有力的基本工具．事实上，有知识的人已经学会问“是这样的概率有多大?”而不问“是这样的吗?”

本书是一本概率论的入门教材，适用于具有初等微积分必备知识的数学、统计学、工程技术以及其他科学（包括计算机科学、社会科学和管理科学）的学生，本书不仅介绍了概率的数学理论，而且通过大量的例子来说明它的许多不同的应用．

第 1 章介绍组合分析的基本原理，这在计算概率时很有用．

第 2 章研究概率论的公理，并说明如何应用这些公理计算各种有趣的概率．

第 3 章论述条件概率与事件独立性的一些极其重要的主题．通过一系列的例题说明：在只有部分信息可利用时，条件概率如何发挥作用；而且即使没有部分信息可利用时，条件概率作为一种工具也可以使我们能比较容易地算出概率．借助“设置条件”获得概率这一极为重要的技巧在第 7 章中会重新提到，在那里我们用它来计算数学期望．

第 4、5、6 章介绍随机变量的概念．第 4 章介绍离散型随机变量，第 5 章介绍连续型随机变量，随机变量的联合分布在第 6 章讨论．随机变量的数学期望和方差等重要概念在第 4 章和第 5 章中介绍，然后对很多常见类型的随机变量确定这两个量的值．

第 7 章介绍数学期望的其他性质．列举许多例题来说明随机变量和的数学期望等于它们的数学期望之和，这一结果是非常有用的．有关条件期望的几节，包括它在预测中的运用和矩母函数都在这一章．另外，在最后一节还介绍了多元正态分布并给出了一个关于多元正态分布的样本均值和样本方差的联合分布的简单证明．

第 8 章阐述概率论中主要的理论结果．特别证明强大数定律和中心极限定理，其中对强大数定律的证明相对简单，即假设随机变量有有限的四阶矩，而对中心极限定理的证明是建立在莱维连续性定理的假设基础之上的．另外，本章还给出诸如马尔可夫不等式、切比雪夫不等式、车尔诺夫界等概率不等式．第 8 章的最后一节给出一个误差的界，此误差为独立的伯努利随机变量之和的概率由具有相同期望值的泊松随机变量的相应概率逼近时的误差．

第 9 章介绍一些补充主题，例如马尔可夫链、泊松过程、信息和编码论．第 10 章讨论模拟．

第 6 版继续对本书的内容进行了改进和调整，添加了许多新练习和例题．添加的例题包括有效性例题（第 4 章的例 4c)、正态逼近例题（第 5 章的例 4i)、对数正态分布在金融中的应用例题（第 6 章的例 3d)，以及不等概率的票券收集例题（第 7 章中的例 2v)．第 7 章还增加了几小节，讨论概率方法（7.2.1 节）和最大-最小恒等式（7.2.2 节)．

与前一版一样，每章最后都有三组习题．它们分别为“习题”、“理论练习”和“自测题与练习”，最后一组习题的详细答案在附录 B 中，这样设计是为了帮助学生测试理解能力和学

习效果.

包括前几版的“概率模型”软盘在内的所有材料均可以从 Ross 网站上下载，网址为 http://cwx. prenhall. com/bookbind/pubbooks/ross. 使用这个网站，可以让学生们在以下六个重要内容上快速容易地计算和模拟：

- 其中的三个模块可以分别导出二项随机变量、泊松随机变量和正态随机变量的概率.
- 另一模块解释了中心极限定理. 它考虑取值为 0，1，2，3，4 之一的随机变量，并允许用户输入这些值的概率和一个数 n，模块则画出这种类型的 n 个独立随机变量之和的概率质量函数. 增大 n 值，可以“看到”质量函数收敛于一个正态密度函数.
- 另外两个模块解释了强大数定律. 同样，用户输入随机变量的 5 个可能值的概率以及一个整数 n，程序则用随机数来模拟具有指定分布的 n 个随机变量. 模块画出每一个结果发生的次数和所有结果发生的平均次数的图. 各个模块对试验结果如何画图是不同的.

我们非常感谢以下审阅者对此书的最近几版提出的有价值的意见和建议：Anastasia Ivanova，北卡罗来纳大学；Richard Bass，康涅狄格大学；Ed Wheeler，田纳西大学；Jean Cadet，纽约州立大学 SUNY 分校；Stony Brook；Jim Propp，威斯康星大学；Mike Hardy，麻省理工学院；Anant Godbole，密歇根技术大学；Zakkula Govindarajulu，肯塔基大学；Richard Groeneveld，艾奥瓦州立大学；Bernard Harris，威斯康星大学；Stephen Herschkorn，拉特格大学；Robert Keener，密歇根大学；Thomas Liggett，加利福尼亚大学洛杉矶分校；Bill McCormick，佐治亚大学；Kathryn Prewitt，亚利桑那州立大学. 特别感谢 Hossein Hamedani（马科萨斯大学）和 Ben Perles 对原稿进行认真核对.

在此还要对早期版本的审阅者表示感谢：Thomas R. Fischer，得克萨斯 A & M 大学；Jay DeVore，加州工业大学 San Luis Obispo 分校；Robb J. Muirhead，密歇根大学；David Heath，康奈尔大学；Myra Samuels，普度大学；I. R. Savage，耶鲁大学；R. Miller，斯坦福大学；K. B. Athreya，艾奥瓦州立大学；Phillip Beckwith，密歇根科技大学；Howard Bird，圣克劳德州立大学；Steven Chiappari，圣克拉大学；James Clay，亚利桑那大学图森分校；Francis Conlan，圣克拉大学；Fred Leysieffer，佛罗里达州立大学；Ian McKeague，佛罗里达州立大学；Helmut Mayer，佐治亚大学；N. U. Prabhu，康奈尔大学；Art Schwartz，密歇根大学安拉伯分校；Therese Shelton，西南大学；Allen Webster，布雷德利大学.

目　录

第1章 组合分析

1.1 引言

这是一个典型的涉及概率的有趣问题．一个通信系统由 n 个看似一样的天线组成，这些天线要以线性顺序排列．合成系统可以接收所有的信号(称为实用的)，只要没有两个相连的天线发生故障．如果结果表明 n 个天线中有 m 个是有缺陷的，那么合成系统实用的概率有多大？例如，令 $n=4$，$m=2$，则有 6 种可能的构形，即

$$
\begin{array}{cccc}
0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 \\
1 & 1 & 0 & 0
\end{array}
$$

其中 1 表示天线正常，0 表示天线有缺陷．由于合成系统在前三种排列中是实用的，而在后三种排列中不是实用的，所以取 $3/6=1/2$ 作为要求的概率似乎是合理的．对于一般的 n，m，也可以用同样的方法计算系统实用的概率，即计算出系统中结果实用的构形数，然后除以所有可能的构形数．

从前面可以看到，找到一个计算事件发生方式的数目的有效方法是非常有用的．事实上，概率论中的很多问题都可以简单地通过计算某特定事件能以几种不同方式发生的数目得以解决．这种计算的数学理论通称为组合分析．

1.2 计数基本原理

下面的计数原理是我们所有工作的基础．简单来说，如果一个试验可以产生 m 个可能结果，而另一个试验可以产生 n 个可能结果，那么这两个试验就有 mn 个可能结果．

计数基本原理 *假设有两个试验，若试验 1 可以产生 m 个可能结果，而对于试验 1 的每一个可能结果，试验 2 都有 n 个可能结果，则这两个试验一共有 mn 个可能结果．*

基本原理的证明 为得到此原理的证明，现将两个试验的全部可能结果列举如下：

$$
\begin{gathered}
(1,1),(1,2),\cdots,(1,n) \\
(2,1),(2,2),\cdots,(2,n) \\
\vdots \\
(m,1),(m,2),\cdots,(m,n)
\end{gathered}
$$

其中结果(i,j)表示试验 1 得到它的第 i 个可能结果，试验 2 得到它的第 j 个可能结果．因此，两个试验的可能结果的集合由 m 行组成，每行含 n 个元素．故结论得证． ■

例 2a 一个小社区有 10 个妇女，每个妇女有 3 个孩子，如果其中一个妇女和她其中的一个孩子被选为本年的母子，那么有多少种不同的选择？

解 把对妇女的选择作为试验 1 的结果，随后对她的孩子的选择作为试验 2 的结果，由基本原理有 $10\times3=30$ 种可能的结果． ■

其中若进行两个以上的试验，基本原理被推广如下.

推广的计数基本原理 如果做 r 个试验，第一个试验有 n_1 个可能结果，对其每一个可能结果试验 2 对应有 n_2 个可能结果，若对前两个试验的每一个可能结果，第三个试验对应有 n_3 个可能结果，……，那么这 r 个试验一共有 $n_1 \cdot n_2 \cdots n_r$ 个可能结果.

2

例 2b 某学院计划委员会由 3 名新生、4 名二年级学生、5 名三年级学生、2 名四年级学生组成，现从该委员会中选各年级 1 人组成 4 人小组委员会，问有多少种不同的选法？

解 我们可以把小组委员会的选择作为从每个年级抽取一名代表的四个独立试验的组合结果，从而由推广的计数基本原理有 $3\times4\times5\times2=120$ 种选法. ■

例 2c 如果前 3 个位置用字母、后 4 个位置用数字，那么有多少个不同的 7 位牌照？

解 由推广的计数基本原理，答案是 $26\cdot26\cdot26\cdot10\cdot10\cdot10\cdot10=175\,760\,000$. ■

例 2d 如果每一个函数值都取 0 或 1，那么定义在 n 个点上的函数有多少个？

解 设这 n 个点分别是 1，2，…，n，因为对每一个 $i=1, 2, \cdots, n$，$f(i)$都等于 0 或者 1，所以有 2^n 种可能的函数. ■

例 2e 在例 2c 中，若字母和数字都不允许重复，那么有多少个牌照？

解 有 $26\cdot25\cdot24\cdot10\cdot9\cdot8\cdot7=78\,624\,000$ 个牌照. ■

1.3 排列

将字母 a，b，c 排成有次序的一列，问有多少种不同的排列？由直接列举法可知有 6 种，即 abc，acb，bac，bca，cab，cba，我们把其中的每一个称为一个排列. 于是包括 3 个对象的集合有 6 种可能的排列. 此结论也可以由计数基本原理推出，因为排列的第一个对象可以是 3 个对象中的任何一个，排列的第二个对象可以从余下的 2 个对象中选出任何一个，而排列的第三个对象只能是剩下的 1 个，因此有 $3\cdot2\cdot1=6$ 种可能排列.

现设有 n 个对象，类似上面的推理可得，这 n 个对象有

$$n(n-1)(n-2)\cdots3\cdot2\cdot1=n!$$

3

种不同的排列.

例 3a 由 9 人组成的棒球队有多少种击球次序？

解 有 $9!=362\,880$ 种可能的击球次序. ■

例 3b 某概率论班共有 6 名男生与 4 名女生，举行一次测验后将他们按成绩排队. 假定他们的分数各不相同，试问：

(a) 可能有多少种排法？

(b) 如果将男、女生分开各自排队，可能有多少种不同的排法？

解 (a) 因为每一队都对应一个包括 10 个人的有次序的排列，从而此题的答案为 $10!=3\,628\,800$.

(b) 因为男生有 6! 种排列，女生有 4! 种排列，从而由基本计数原理得可能有 $(6!)(4!)=(720)(24)=17\,280$ 种排法. ■

例 3c 琼斯有 10 本书，其中数学书 4 本，化学书 3 本，历史书 2 本，语文书 1 本，琼斯打算把他的书摆在书架上，若要求同类的书放在一起，问可能有多少种不同的摆法？

解 把数学书摆在前面，接着是化学书，然后是历史书，最后是语文书，共有 4! 3! 2! 1! 种摆法. 类似地，对每一种书类的次序都有 4! 3! 2! 1! 种摆法，由于书类又有 4! 种次序，

故所求的答案是4! 4! 3! 2! 1! =6 912. ■

现在我们来计算当有一些对象彼此不加区别时，含有 n 个对象的集合的排列数. 为直观起见，考虑如下例子.

例 3d 用字母 *PEPPER* 可以组成多少种不同的字母排列?

解 首先我们指出，若3个 P 之间和2个 E 之间彼此加以区别，则字母 $P_1E_1P_2P_3E_2R$ 共有6! 种排列. 然而，考虑这些排列中的任何一个，例如 $P_1P_2E_1P_3E_2R$，若在 P 与 P 之间、E 与 E 之间交换位置，则所得的排列仍将是 *PPEPER* 型的，也就是说，3! 2! 个排列

4

$$
\begin{array}{ll}
P_1P_2E_1P_3E_2R & P_1P_2E_2P_3E_1R \\
P_1P_3E_1P_2E_2R & P_1P_3E_2P_2E_1R \\
P_2P_1E_1P_3E_2R & P_2P_1E_2P_3E_1R \\
P_2P_3E_1P_1E_2R & P_2P_3E_2P_1E_1R \\
P_3P_1E_1P_2E_2R & P_3P_1E_2P_2E_1R \\
P_3P_2E_1P_1E_2R & P_3P_2E_2P_1E_1R
\end{array}
$$

都是 *PPEPER* 型的. 因此，字母 *PEPPER* 有6! /(3! 2!)=60 种可能的排列. ■

一般地，运用例3d的推理可证：若 n 个对象中有 n_1 个对象彼此不加区别，另 n_2 个对象彼此不加区别，……，另 n_r 个对象彼此不加区别，则总共有

$$\frac{n!}{n_1!\ n_2!\ \cdots n_r!}$$

种不同排列.

例 3e 一次国际象棋比赛有10名选手，其中4名俄罗斯人，3名美国人，2名英国人，1名巴西人，如果比赛结果的次序正好是选手国籍的次序，那么有多少种可能的结果?

解 有 $\dfrac{10!}{4!\ 3!\ 2!\ 1!}=12\,600$ 种可能的结果. ■

例 3f 一组旗由4面白旗、3面红旗、2面蓝旗组成，并认为颜色相同的旗子是一样的，那么这9面旗子排成一行可以形成多少种不同的记号?

解 有 $\dfrac{9!}{4!\ 3!\ 2!}=1\,260$ 种不同的记号. ■

1.4 组合

我们经常要计算从 n 个对象中取出的、包括 r 个对象的不同组的数目. 例如，从 A，B，C，D 和 E 这五项中取出三项为一组，可能有多少不同的组? 为回答这个问题，可进行如下推理：由于选择第一项有5种方法，然后选第二项有4种方法，再选第三项有3种方法，故若考虑项目被选取的次序，则选取含三项的组有 $5\cdot4\cdot3$ 种方法. 但此时每个由某三项组成的组，例如由 A，B 和 C 组成的组，都被计数6次(也就是说，若考虑选取次序，排列 ABC，ACB，BAC，BCA，CAB，CBA 全要算上)，由此可见，总共能组成

5

$$\frac{5\cdot4\cdot3}{3\cdot2\cdot1}=10$$

个不同的组.

一般地，由于考虑选取次序时，从 n 项中选取 r 项为一组，共有 $n(n-1)\cdots(n-r+1)$ 种不同选法，而这时每个含 r 项的组全被计数 $r!$ 次，故由一个包含 n 项的集合中取出 r 项所组成

的不同的组数应是

$$\frac{n(n-1)\cdots(n-r+1)}{r!}=\frac{n!}{(n-r)!\ r!}$$

记号与术语 对 $r\leqslant n$，我们定义

$$\binom{n}{r}=\frac{n!}{(n-r)!\ r!}$$

并称 $\binom{n}{r}$ 为从 n 个对象中一次取出 r 个的可能组合数. [1]

于是，$\binom{n}{r}$ 表示不考虑选取次序时，从一个包含 n 个对象的集合中取出的、包括 r 个对象的不同组的数目.

例 4a 从 20 个人中选 3 人组成一个委员会，问有多少种选法？

6

解 有 $\binom{20}{3}=\frac{20\cdot 19\cdot 18}{3\cdot 2\cdot 1}=1\,140$ 种可能选法. ■

例 4b 从 5 个女人与 7 个男人中选 2 女 3 男组成一个委员会，有多少种不同的选法？如果其中有 2 个男人长期不和，不能同时进入委员会，那么又有多少种选法？

解 由于 2 个女人有 $\binom{5}{2}$ 种选法，3 个男人有 $\binom{7}{3}$ 种选法，从而由计数基本原理知由 2 女 3 男组成的委员会共有

$$\binom{5}{2}\binom{7}{3}=\left(\frac{5\cdot 4}{2\cdot 1}\right)\frac{7\cdot 6\cdot 5}{3\cdot 2\cdot 1}=350$$

种选法.

另外，如果其中 2 个男人拒绝一起进入委员会，那么 3 个男人中不包括 2 个不和男人中的任意一个有 $\binom{2}{0}\binom{5}{3}$ 种选法，3 个男人中仅包括 2 个不和男人中的一个有 $\binom{2}{1}\binom{5}{2}$ 种选法，这就是说不同时包括两个不和男人的 3 个男人有 $\binom{2}{0}\binom{5}{3}+\binom{2}{1}\binom{5}{2}=30$ 种选法. 因为 2 个女人有 $\binom{5}{2}$ 种选法，因此这种情况下有 $30\binom{5}{2}=300$ 种不同的选法. ■

例 4c 考虑 n 个天线，其中 m 个天线有缺陷，$n-m$ 个天线是实用的，假设所有有缺陷的天线和所有实用的天线没有区别，那么任意两个有缺陷的天线不连接的接线方法有多少种？

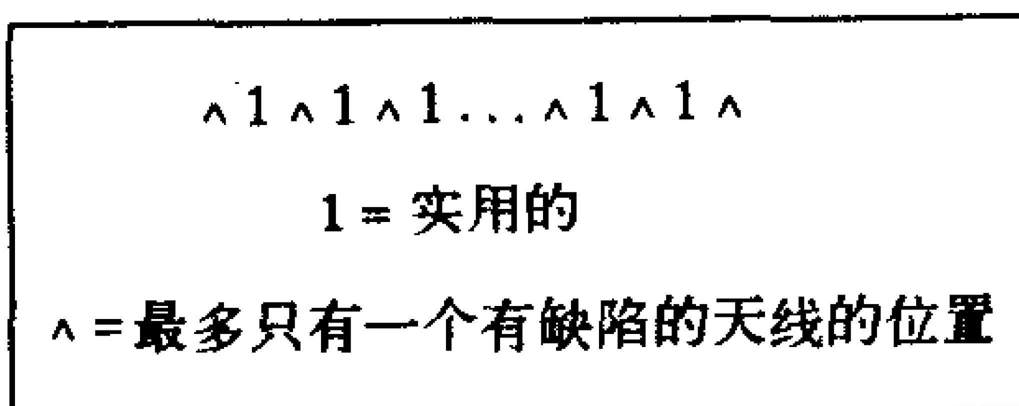


图 1-1

解 对 $n-m$ 个实用的天线排队，现在假设任意两个有缺陷的天线均不连接，那么两个实用天线之间最多只有一个有缺陷的天线，这就是说在 $n-m$ 个实用天线之间的 $n-m+1$ 个位置中——图 1-1 中用脱字符号表示，必须选择其中的 m 个位置

[1] 习惯上，约定 0! 为 1，于是 $\binom{n}{0}=\binom{n}{n}=1$. 当 $i<0$ 或 $i>n$ 时，我们取 $\binom{n}{i}$ 等于 0.

来放有缺陷的天线，从而有$\binom{n-m+1}{m}$种连接，其中在任意两个有缺陷的天线之间至少有一个实用的天线. ■ 7

一个有用的组合恒等式为

$$\binom{n}{r}=\binom{n-1}{r-1}+\binom{n-1}{r} \quad 1\leqslant r\leqslant n \tag{4-1}$$

式(4-1)可以用解析的方法证明，也可用如下组合方法证明. 考虑一个包含 n 个对象的集合并挑出其中特定的某一个——称之为 1 号对象. 于是，包含 1 号对象且容量为 r 的组合数为$\binom{n-1}{r-1}$(因为每一个这样的组合都是从剩下的 $n-1$ 个对象中再取 $r-1$ 个组成). 同样地，不含 1 号对象且容量为 r 的组合数为$\binom{n-1}{r}$. 由于容量为 r 的组合总数为$\binom{n}{r}$，式(4-1)得证.

通常称$\binom{n}{r}$为**二项式系数**，这是由于它在二项式定理中占有重要的地位.

二项式定理

$$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} \tag{4-2}$$

我们将给出二项式定理的两个证明，前者用数学归纳法，而后者则基于组合法.

二项式定理的归纳法证明 当 $n=1$ 时，式(4-2)化为

$$x+y=\binom{1}{0}x^0y^1+\binom{1}{1}x^1y^0=y+x$$

设式(4-2)对 $n-1$ 成立，那么

$$(x+y)^n=(x+y)(x+y)^{n-1}=(x+y)\sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-1-k}$$

$$=\sum_{k=0}^{n-1}\binom{n-1}{k}x^{k+1}y^{n-1-k}+\sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-k}$$

8

在第一个和式中令 $i=k+1$，在第二个和式中令 $i=k$，则有

$$(x+y)^n=\sum_{i=1}^{n}\binom{n-1}{i-1}x^i y^{n-i}+\sum_{i=0}^{n-1}\binom{n-1}{i}x^i y^{n-i}$$

$$=x^n+\sum_{i=1}^{n-1}\left[\binom{n-1}{i-1}+\binom{n-1}{i}\right]x^i y^{n-i}+y^n$$

$$=x^n+\sum_{i=1}^{n-1}\binom{n}{i}x^i y^{n-i}+y^n=\sum_{i=0}^{n}\binom{n}{i}x^i y^{n-i}$$

其中倒数第二个等式是由式(4-1)得到的. 由归纳法，定理得证. ■

二项式定理的组合法证明 考虑乘积

$$(x_1+y_1)(x_2+y_2)\cdots(x_n+y_n)$$

它的展开式是 2^n 项的和，每一项是 n 个因子的积，并且对每一个 $i=1, 2\cdots, n$，和式中 2^n 项的各项都将包含 x_i 或 y_i. 例如，

$$(x_1+y_1)(x_2+y_2)=x_1x_2+x_1y_2+y_1x_2+y_1y_2$$

现问：在和式这 2^n 项中恰由 k 个 x_i 因子、$(n-k)$个 y_i 因子组成的项有多少？由于 k 个 x_i 与 $(n-k)$个 y_i 组成的每一项正好对应着从 n 个数 $x_1,x_2,\cdots,x_n$ 中任选 k 个的一种选法，故这样的项应有$\binom{n}{k}$个. 令 $x_i=x,y_i=y,i=1,\cdots,n$，于是得到

$$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$$ ■

例 4d 展开$(x+y)^3$.

解

$$(x+y)^3=\binom{3}{0}x^0y^3+\binom{3}{1}x^1y^2+\binom{3}{2}x^2y^1+\binom{3}{3}x^3y^0=y^3+3xy^2+3x^2y+x^3$$ ■

例 4e n 个元素的集合总共有多少个子集？

解 因为容量为 k 的子集共$\binom{n}{k}$个，故要求的答案是

9

$$\sum_{k=0}^{n}\binom{n}{k}=(1+1)^n=2^n$$

这个结果也可由下述方法得到：把 0 或 1 中的一个数分配给此集合中的每一个元素. 若以全体分配到 1 的元素构成一个子集，则分配方法与子集之间是一一对应的. 由于总共有 2^n 种可能的分配方法，故子集总数为 2^n.

注意到已经把零个元素组成的集合(即空集)算作一个子集，所以至少含一个元素的子集数应为 2^n-1. ■

1.5 多项式系数

本节考虑下列问题：将含有 n 个元素的集合分为各含 $n_1,n_2,\cdots,n_r$ 个元素的 r 个不同的组，其中 $\sum_{i=1}^{r}n_i=n$，问可能有多少种不同的分法?为回答这一问题，我们注意到，第一组有$\binom{n}{n_1}$种可能选法；对第一组的每一种选法，第二组有$\binom{n-n_1}{n_2}$种可能选法；对前两组的每一种选法，第三组有$\binom{n-n_1-n_2}{n_3}$种可能选法；以此类推. 因此，根据推广的计数基本原理可知，共有

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{r-1}}{n_r}$$
$$=\frac{n!}{(n-n_1)!\ n_1!}\frac{(n-n_1)!}{(n-n_1-n_2)!\ n_2!}\cdots\frac{(n-n_1-n_2-\cdots-n_{r-1})!}{0!\ n_r!}=\frac{n!}{n_1!\ n_2!\ \cdots\ n_r!}$$

种可能的分法.

记号 若 $n_1+n_2+\cdots+n_r=n$，我们定义

$$\binom{n}{n_1,n_2,\cdots,n_r}=\frac{n!}{n_1!\ n_2!\ \cdots\ n_r!}$$

于是，若将 n 个不同的对象分成容量各为 $n_1,n_2,\cdots,n_r$ 的 r 个不同的组，则$\binom{n}{n_1,n_2,\cdots,n_r}$表示

可能分法的数目.

10

例 5a 某小城市的一个警察局有 10 名警官，如果需要 5 名警官去巡逻、2 名警官专门在局机关值班、3 名警官留在局里作为后备，问把这 10 名警官分成 3 组可能有多少种不同分法？

解 有$\frac{10!}{5!\ 2!\ 3!}=2\,520$种可能分法. ■

例 5b 将 10 个孩子分为各 5 人的 A，B 两队，A 队去参加某一场比赛，B 队则去参加另一场比赛. 问可能有多少种不同分法？

解 有$\frac{10!}{5!\ 5!}=252$种可能分法. ■

例 5c 10 个孩子要举行一场篮球赛. 他们在操场上分成各 5 人的两队，问可能有多少种不同分法？

解 注意此例与例 5b 不同，因为现在不考虑两队的次序. 也就是说，只要分成两队，不存在 A 队 B 队的问题. 因此，所求的答案是

$$\frac{10!\ /(5!\ 5!)}{2!}=126$$ ■

下面的定理是二项式定理的推广，其证明留作练习.

多项式定理

$$(x_1+x_2+\cdots+x_r)^n=\sum_{\substack{(n_1,\cdots,n_r):\\ n_1+n_2+\cdots+n_r=n}}\binom{n}{n_1,n_2,\cdots,n_r}x_1^{n_1}x_2^{n_2}\cdots x_r^{n_r}$$

也就是说，求和是取遍一切满足 $n_1+n_2+\cdots+n_r=n$ 的非负整值向量$(n_1,n_2,\cdots,n_r)$.

11

$\binom{n}{n_1,n_2,\cdots,n_r}$通常称为多项式系数.

例 5d

$$\begin{aligned}(x_1+x_2+x_3)^2&=\binom{2}{2,0,0}x_1^2x_2^0x_3^0+\binom{2}{0,2,0}x_1^0x_2^2x_3^0+\binom{2}{0,0,2}x_1^0x_2^0x_3^2\\&\quad+\binom{2}{1,1,0}x_1^1x_2^1x_3^0+\binom{2}{1,0,1}x_1^1x_2^0x_3^1+\binom{2}{0,1,1}x_1^0x_2^1x_3^1\\&=x_1^2+x_2^2+x_3^2+2x_1x_2+2x_1x_3+2x_2x_3\end{aligned}$$ ■

*1.6 方程整数解的个数[⊖]

将 n 个不同的球分放到 r 个不同的箱中，共有 r^n 个可能结果. 这是因为每个球都可能被放入这 r 个箱中的任何一个. 但是，如果假定 n 个球彼此不加区别，此时又有多少种不同结果呢？由于球不加区别，故可以把分 n 个球到 r 个箱中这一试验的结果看作一个向量$(x_1,x_2,\cdots,x_r)$，其中 x_i 表示被分到第 i 个箱中的球数，因此，问题化为求满足

$$x_1+x_2+\cdots+x_r=n$$

的各非负整值向量$(x_1,x_2,\cdots,x_r)$的个数. 为了解决这个问题，我们先考虑正整数解的个数，

⊖ “*”号表示此内容为选学内容.

为此假设对 n 个不同的物体排队，并把它们分到 r 个非空的组中，从相邻物体之间的 $n-1$ 个空中选择 $r-1$ 个作为分割点(见图 1-2)．例如，令 $n=8$，$r=3$，选取如下所示的 2 个分割点：

$$000|000|00$$

从而得到的向量为 $x_1=3$，$x_2=3$，$x_3=2$，因为有 $\binom{n-1}{r-1}$ 个可能的结果，所以得到以下命题．

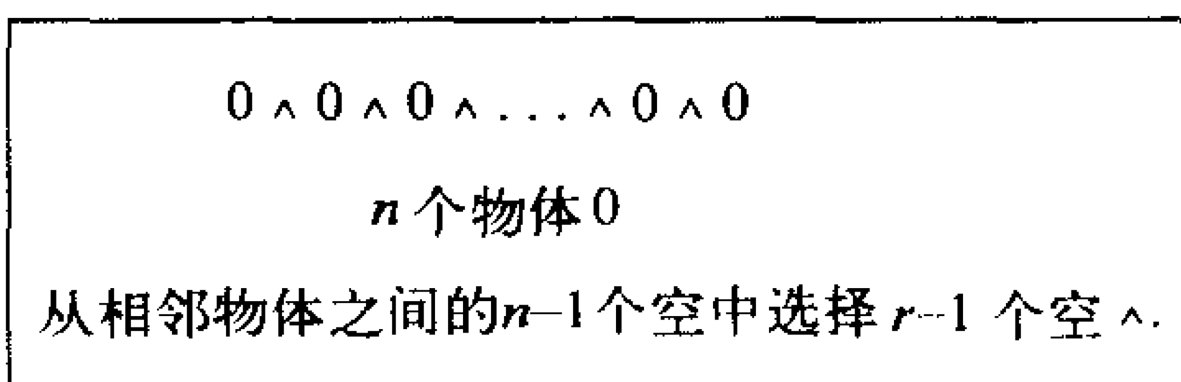


图 1-2

命题 6.1 满足条件

$$x_1+x_2+\cdots+x_r=n(x_i>0, i=1,\cdots,r)$$

的正整值向量$(x_1,x_2,\cdots,x_r)$有$\binom{n-1}{r-1}$个．

为了得到非负(相对于正的)解的个数，注意到方程 $x_1+x_2+\cdots+x_r=n$ 非负解的个数与方程 $y_1+\cdots+y_r=n+r$ 正数解的个数相同(令 $y_i=x_i+1$，$i=1,\cdots,r$)，因此，由命题 6.1 得到以下命题．

命题 6.2 满足条件

$$x_1+x_2+\cdots+x_r=n \tag{6-1}$$

的非负整值向量$(x_1,x_2,\cdots,x_r)$有$\binom{n+r-1}{r-1}$个．

例 6a $x_1+x_2=3$ 有多少组互不相同的非负整数解？

解 有$\binom{3+2-1}{2-1}=4$ 组解，即$(0,3),(1,2),(2,1),(3,0)$． ■

例 6b 一位投资者要在 4 个投资项目上投资 2 万美元，每一个投资项目必须以 1 千美元为单位投资．如果 2 万美元全部投资，那么有多少种不同的投资策略？如果不需要所有的钱都投资又有多少种投资策略？

解 如果令 $x_i(i=1,2,3,4)$表示投资在第 i 个项目上的钱数(以千计)，那么当所有的钱都投资时，x_1,x_2,x_3,x_4 是满足 $x_1+x_2+x_3+x_4=20(x_i\geqslant 0)$的整数．因此，由命题 6.2 有 $\binom{23}{3}=1\ 771$种投资策略．如果并非所有的钱都投资，那么令 x_5 表示没有投资的钱数，一个策略是满足 $x_1+x_2+x_3+x_4+x_5=20$ 的非负整值向量(x_1,x_2,x_3,x_4,x_5)，因此由命题 6.2 有 $\binom{24}{4}=10\ 626$种投资策略． ■

12 ~ 13

例 6c $(x_1+x_2+\cdots+x_r)^n$ 的多项展开式中共有多少项？

解

$$(x_1+x_2+\cdots+x_r)^n=\sum\binom{n}{n_1,\cdots,n_r}x_1^{n_1}\cdots x_r^{n_r}$$

其中，求和是取遍满足 $n_1+\cdots+n_r=n$ 的所有非负整值$(n_1,\cdots,n_r)$，故由命题 6.2 知，此多项展开式共有$\binom{n+r-1}{r-1}$项． ■

例 6d 考虑例 4c，其中有一个 n 项的集合，其中 m 项(无区别)有缺陷，余下的 $n-m$ 项(也是无区别)实用，我们的目标是找到任意两个有缺陷的天线不连接的接线方法有多少种．为了确定这个量，假设有缺陷的项自身进行排列，实用的项再进行排列，假设 x_1 表示放在第一个

有缺陷项左边的实用项的个数，x_2 表示放在前两个有缺陷项之间的实用项的个数，如此下去. 这样，就有下面的图解：

$$x_1\ 0\ x_2\ 0\cdots x_m\ 0\ x_{m+1}$$

现在只要 $x_i>0(i=2,\cdots,m)$，则任意一对有缺陷项之间就至少有一个实用项，因此满足条件的结果的数目就是满足下列条件的向量 $x_1,\cdots,x_{m+1}$ 的个数：

$$x_1+\cdots+x_{m+1}=n-m \quad x_1\geqslant 0, x_{m+1}\geqslant 0, x_i>0, i=2,\cdots,m$$

但是，假设令 $y_1=x_1+1, y_i=x_i, i=2,\cdots,m, y_{m+1}=x_{m+1}+1$，可以看到其结果和满足下列条件的正向量$(y_1,\cdots,y_{m+1})$的个数相同：

$$y_1+y_2+\cdots+y_{m+1}=n-m+2$$

因此，由命题 6.1 有 $\binom{n-m+1}{m}$ 个结果，这和例 4c 的结果是一样的.

假设现在我们对每一对有缺陷项都被至少 2 个实用项分开的结果的数目感兴趣，用上述同样的方法，其结果等于满足下列条件的向量个数：

14

$$x_1+\cdots+x_{m+1}=n-m \quad x_1\geqslant 0, x_{m+1}\geqslant 0, x_i\geqslant 2, i=2,\cdots,m$$

令 $y_1=x_1+1, y_i=x_i-1, i=2,\cdots,m, y_{m+1}=x_{m+1}+1$，则其结果和下列方程的正解的个数相同：

$$y_1+\cdots+y_{m+1}=n-2m+3$$

因此，由命题 6.1 有 $\binom{n-2m+2}{m}$ 个结果. ■

小结

如果一个试验由两部分组成，第一部分有 n 个可能结果，对其每一个结果第二部分有 m 个可能结果，那么此试验有 nm 个可能结果，这就是计数基本原理.

n 项有 $n!=n(n-1)\cdots 3\cdot 2\cdot 1$ 种线性排列，并且定义 $0!=1$.

令 $\binom{n}{i}=\frac{n!}{(n-i)!\ i!}$，其中 $0\leqslant i\leqslant n$，对于其他的 i 令它为 0. 这个量表示从一个含有 n 项的集合中抽取含有 i 项的不同子集的个数，因为它在二项式定理中非常重要，所以通常称为二项式系数，即

$$(x+y)^n=\sum_{i=0}^{n}\binom{n}{i}x^iy^{n-i}$$

对于和为 n 的非负整数 n_1，…，n_r，

$$\binom{n}{n_1,n_2,\cdots,n_r}=\frac{n!}{n_1!\ n_2!\ \cdots n_r!}$$

表示将 n 个元素分为 r 个独立的不互相重叠的组的分法数，这些组分别包括 $n_1,n_2,\cdots,n_r$ 个独立元素.

习题

1. (a) 如果前两位用字母、后 5 位用数字，问可能有多少个不同的 7 位牌照?
   (b) 若进一步要求同一牌照所用的字母与数字都不重复，那么可能有多少个 7 位牌照?
2. 一个骰子掷 4 次可能有多少种结果? 例如，第一次掷 3 点，第二次掷 4 点，第三次掷 3 点，第四次掷 1

点，那么结果就是 3，4，3，1.

[15]

3. 分配 20 名工人做 20 种不同的工作，有多少种分配方式？

4. 约翰、吉姆、杰伊和杰克组成一个有 4 种乐器的小乐队，如果每个人都会演奏这 4 种乐器，问可能有多少种不同排列？如果约翰和吉姆会演奏这 4 种乐器，而杰伊和杰克只会弹钢琴和打鼓．那么可能有多少种排列？

5. 许久以前，美国和加拿大的电话地区码是由 3 个数字组成的序列，第一个数字是 2～9 之间的一个整数，第二个数字是 0 或 1，第三个数字是 1～9 之间的任意整数，问有多少种地区码？又有多少种以 4 开始的地区码？

6. 有一首著名的儿歌是这样的：

   当我去 Ives 街
   碰到了一个有 7 个妻子的男士
   每一个妻子有 7 只袋子
   每一个袋子有 7 只猫
   每一只猫有 7 只小猫

   这个旅行者看到了多少只小猫？

7. (a) 3 个男孩 3 个女孩坐在一排，有多少种坐法？
   (b) 若进一步要求男孩和女孩分别坐在一起，有多少种坐法？
   (c) 若只有男孩必须坐在一起，有多少种坐法？
   (d) 如果相邻的座位上必须坐不同性别的孩子，问有多少种坐法？

8. 由下列各组字母能组成多少种不同的字母排列？
   (a) FLUKE
   (b) PROPOSE
   (c) MISSISSIPPI
   (d) ARRANGE

9. 一个孩子有 12 块积木，其中 6 块黑的、4 块红的、1 块白的、1 块蓝的，孩子想把积木排成一行，问有多少种排法？

10. 8 个人坐成一排，问有多少种坐法？
    (a) 在座位安排上没有限制.
    (b) A 和 B 必须坐在一起.
    (c) 有 4 位男士、4 位女士，并且任意 2 位男士或 2 位女士不能坐在一起.
    (d) 有 5 位男士，并且他们必须坐在一起.
    (e) 有 4 对夫妇，每一对夫妇必须坐在一起.

11. 3 本小说、2 本数学书、1 本化学书摆在书架上，有多少种排列方法？
    (a) 这些书可以任意排列.
    (b) 数学书必须放在一起，小说必须放在一起.
    (c) 小说必须放在一起，其他书可以任意放.

12. 五个奖(最佳成绩奖、最佳领导奖，等等)分别颁给从一个 30 人的班中选取的学生，有多少种不同的结果？
    (a) 一个学生可以得任意多个奖.
    (b) 每一个学生最多得一个奖.

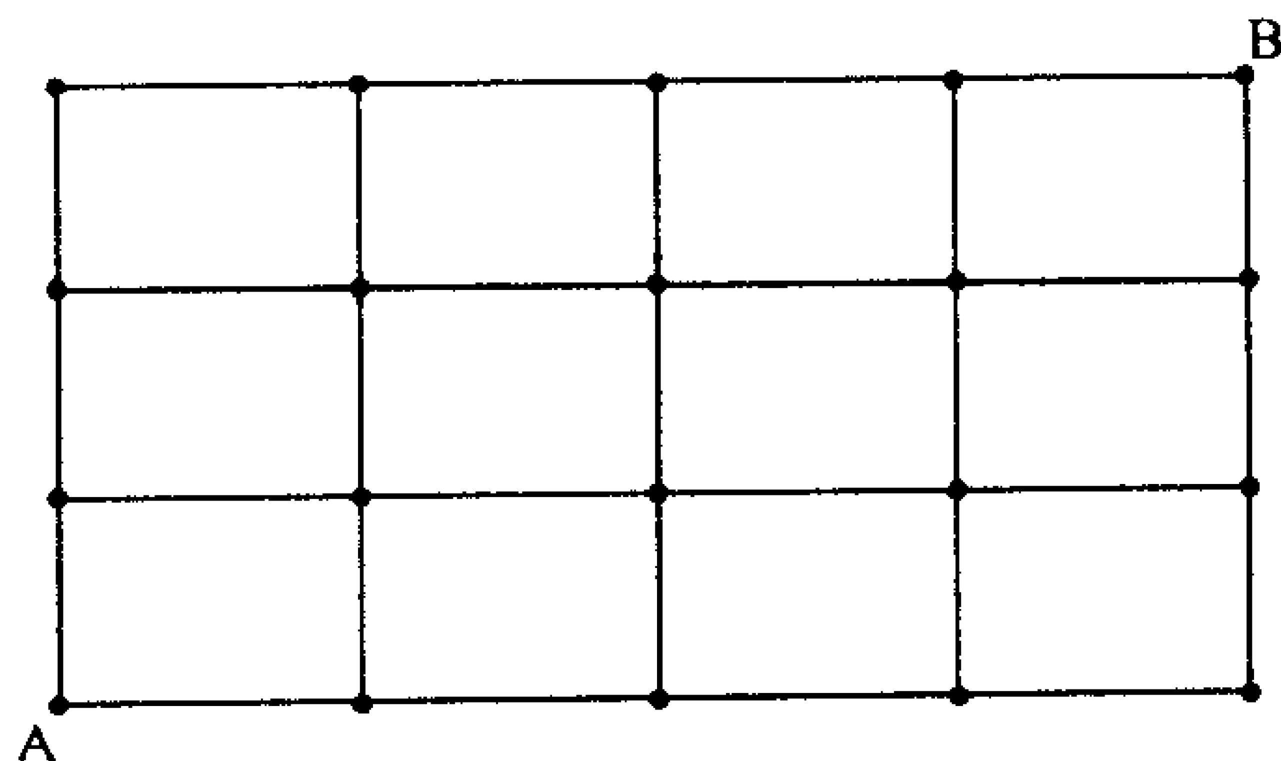
13. 考虑一个 20 人的组，如果每一个人都要和其余的人握手，那么有多少次握手？

14. 从一副扑克牌的 52 张中任取 5 张，有多少种可能结果？

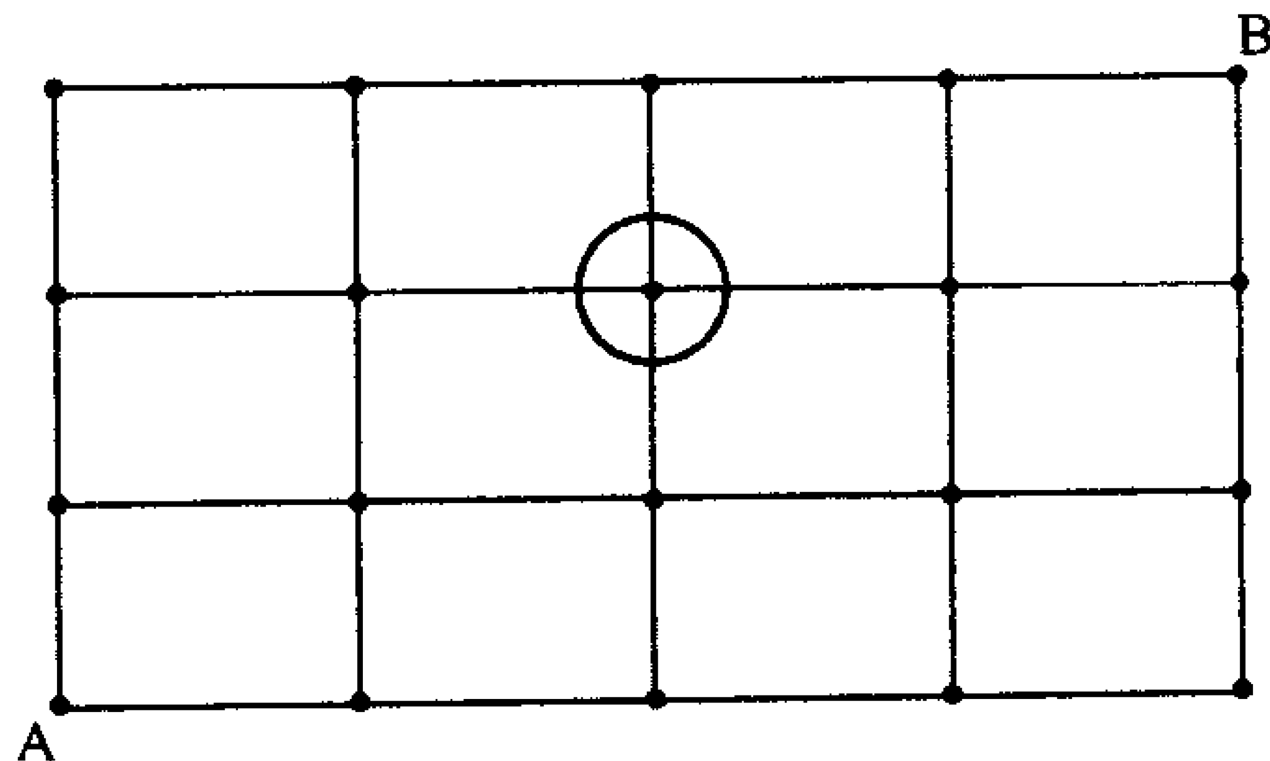
15. 某舞蹈班有 22 名学生，其中 10 名女生、12 名男生，如果选出 5 名男生、5 名女生配成对，有多少种结果？

16

16. 某学生不得不从 6 本数学书、7 本科学书、4 本经济学书中选出 2 本卖掉，有多少种选法？
    (a) 两本书是同一科目.
    (b) 两本书是不同科目.
17. 将 7 个不同的礼物分赠给 10 个孩子，如果每一个孩子最多有一个礼物，那么有多少种不同结果？
18. 从 5 名共和党人、6 名民主党人和 4 名无党派人中选出 2 名共和党人、2 名民主党人和 3 名无党派人组成一个 7 人委员会，问有多少种选法？
19. 从 8 位男士和 6 位女士中选 3 位男士和 3 位女士组成一个委员会，问有多少种选法？
    (a) 其中 2 位男士拒绝一起进入委员会.
    (b) 其中 2 位女士拒绝一起进入委员会.
    (c) 其中 1 位男士和 1 位女士拒绝一起进入委员会.
20. 某人有 8 个朋友，他想邀请其中 5 个参加晚会.
    (a) 如果他的两个朋友由于长期不和不能同来参加晚会，有多少种邀请方案？
    (b) 如果他的两个朋友必须一起参加晚会，又有多少种方案可供选择？
21. 考虑下面由点组成的网格，现在从点 A 出发，每次可以向上或向右移动一步，直到到达 B 点，那么从 A 点到 B 点有多少条不同的路线？
    提示：注意从 A 点到 B 点必须向右走 4 步、向上走 3 步.



22. 在 21 题中，穿过用圆圈圈住的点从 A 点到 B 点有多少条不同的路线？



17

23. 某从事梦幻研究的心理学实验室有 3 个房间，每个房间有 2 个床位，现将 3 对双胞胎分配到这 6 张床上，使得每一对双胞胎正好睡在同一个房间的两个不同的床上，问有多少种分配方案？
24. 展开 $(3x^2+y)^5$.
25. 4 个人打桥牌，每人分配 13 张，问有多少种分法？
26. 展开 $(x_1+2x_2+3x_3)^4$.
27. 现将 12 个人分成各含 3、4、5 人的三组，问有多少种分法？
28. 将 8 名新教师分配到 4 所学校，有多少种分配方案？若分给每所学校两名教师，又有多少种分配

方案?

29. 10 名举重运动员参加一次比赛，其中美国运动员 3 名、俄罗斯运动员 4 名、中国运动员 2 名、加拿大运动员 1 名，如果只按运动员所代表的国家计分，而不计算个人成绩，问可能有多少不同的结果? 美国正好有一名运动员在前三名之中、另两名在后三名之中的不同结果有多少种?
30. 包括俄罗斯、法国、英国和美国在内的 10 个国家各一名代表在一排椅子上就坐，如果法国代表和英国代表必须坐在一起，而俄罗斯代表和美国代表不能挨着坐，问可能有多少种不同坐法?

*31. 如果 8 块相同的黑板分配给 4 个学校，有多少种分法? 如果每个学校必须接收至少一块黑板又有多少种分法?

*32. 电梯载着 8 个人(不包括电梯操作员)在底层起动，到顶层(6 楼)后乘客已全部出去. 如果操作员只注意到每层楼出去几个人(不管是哪几个人)，那么他能看到多少种离开电梯的方式? 如果 8 名乘客中有 5 名男士和 3 名女士，而操作员又注意了分辨男女，问题的答案应是多少?

*33. 现将 2 万美元投资在 4 个可能的项目上，每个投资项目必须以 1 千美元为单位投资，并且如果要在这些项目上投资必须有最小投资，最小投资为 2 千美元、2 千美元、3 千美元、4 千美元，则有多少种不同的投资策略?
   (a) 在每一个项目上都必须投资.
   (b) 至少在 4 个项目中的 3 个项目上投资.

理论练习

1. 证明计数基本原理的推广形式.
2. 进行两个试验，第一个试验有 m 个可能结果，若第一个试验得到第 i 个结果，则第二个试验有 n_i 个可能结果，$i=1,2,\cdots,m$. 问这两个试验的可能结果共有多少个?
3. 如果考虑选取次序，那么从一个包含 n 个对象的集合中选取 r 个对象有多少种不同的选法?

18

4. r 个黑球与 $n-r$ 个白球的线性排列共有 $\binom{n}{r}$ 种，试给出这一事实的组合解释.
5. 求 x_i 取 0 或 1 并满足 $\sum_{i=1}^{n} x_i \geqslant k$ 的向量 $(x_1,\cdots,x_n)$ 的个数.
6. 每个 x_i 都是正整数，且满足 $1\leqslant x_i\leqslant n$ 和 $x_1<x_2<\cdots<x_k$，问有多少这样的向量 $x_1,\cdots,x_k$?
7. 试用解析方法证明式(4-1).
8. 证明

$$\binom{n+m}{r}=\binom{n}{0}\binom{m}{r}+\binom{n}{1}\binom{m}{r-1}+\cdots+\binom{n}{r}\binom{m}{0}$$

   提示：考虑由 n 位男士与 m 位女士组成的人群，从中选出 r 个人为一组有多少种选法?
9. 用理论练习题 8 证明

$$\binom{2n}{n}=\sum_{k=0}^{n}\binom{n}{k}^2$$

10. 从 n 个人中抽取 $k(k\leqslant n)$ 个人组成一个委员会，其中一人被任命为主席.
   (a) 首先选择委员会，然后选择主席，证明有 $\binom{n}{k}k$ 种可能的选法.
   (b) 首先选择非主席委员会成员，然后选择主席，证明有 $\binom{n}{k-1}(n-k+1)$ 种可能的选法.
   (c) 首先选择主席，然后选择其他委员会成员，证明有 $n\binom{n-1}{k-1}$ 种可能的选法.

(d) 从(a)、(b)和(c)证明

$$k\binom{n}{k}=(n-k+1)\binom{n}{k-1}=n\binom{n-1}{k-1}$$

(e) 由$\binom{m}{r}$的阶乘定义证明(d).

11. 下面的恒等式是著名的费马组合恒等式：

$$\binom{n}{k}=\sum_{i=k}^{n}\binom{i-1}{k-1} \qquad n\geqslant k$$

给出一个组合方法(不需要计算)证明这个恒等式.

提示：考虑从 1 到 n 的集合，有多少个含 k 个元素的子集把 i 作为最高号元素？

12. 考虑下列组合恒等式：

$$\sum_{k=1}^{n}k\binom{n}{k}=n\cdot 2^{n-1}$$

19

(a) 考虑含 n 个人的集合，从中选出任意人组成一个委员会，其中一人被任命为主席，试用两种组合方法证明上式.

提示：(i) 对由 k 个人组成的委员会，并且其中一人为主席，有多少种可能的选法？

(ii) 有一个主席和其他委员会成员，有多少种可能的选法？

(b) 对 $n=1,2,3,4,5$，证明

$$\sum_{k=1}^{n}\binom{n}{k}k^2=2^{n-2}n(n+1)$$

为给上式一个组合法证明，考虑一个 n 个人的集合，讨论等式两边各代表对委员会、其主席和秘书(也可能和主席是同一人)的不同的选法.

提示：(i) 委员会恰好有 k 个人，有多少种不同的选法？

(ii) 主席和秘书是一个人，有多少种不同的选法？(答案：$n2^{n-1}$.)

(iii) 主席和秘书不是同一个人，有多少种不同的选法？

(c) 证明

$$\sum_{k=1}^{n}\binom{n}{k}k^3=2^{n-3}n^2(n+3)$$

13. 对所有的 $n>0$，证明

$$\sum_{i=0}^{n}(-1)^i\binom{n}{i}=0$$

提示：使用二项式定理.

14. 从 n 个人中选出 j 人组成一个委员会，再从此委员会中选出 i 人组成一个小组委员会，其中 $i\leqslant j$.

(a) 用两种方法，通过计算委员会和小组委员会的可能选法得到一个组合恒等式——首先假设先选委员会然后选小组委员会，再假设先选小组委员会然后选委员会的其余成员.

(b) 用(a)证明下列组合恒等式：

$$\sum_{j=1}^{n}\binom{n}{j}\binom{j}{i}=\binom{n}{i}2^{n-i} \qquad i\leqslant n$$

(c) 用(a)和理论练习题 13 证明

$$\sum_{j=i}^{n}\binom{n}{j}\binom{j}{i}(-1)^{n-j}=0 \qquad i\leqslant n$$

20

15. 令 $H_k(n)$表示满足 $1\leqslant x_i\leqslant n, x_1\leqslant x_2\leqslant\cdots\leqslant x_k$ 的向量 $x_1,\cdots,x_k$ 的个数，其中每一个 x_i 都是正整数

(a) 不用计算证明

$$H_1(n)=n$$

$$H_k(n) = \sum_{j=1}^{n} H_{k-1}(j) \qquad k > 1$$

提示：$x_k=j$ 的向量有多少个？

(b) 用上述递归式计算 $H_3(5)$.

提示：首先对 $n=1,2,3,4,5$ 计算 $H_2(n)$.

16. 考虑一个 n 个选手的比赛，其结果就是这些选手的名次，允许平局. 也就是说，按其结果将选手分组，第一组是那些与第一名平局的选手，第二组是那些与第二名平局的选手，以此类推. 令 $N(n)$表示不同的结果数，例如，$N(2)=3$，这是因为一场比赛中有 2 名选手，选手 1 可以是唯一的第一名，选手 2 也可以是唯一的第一名，或者与选手 1 平局.

(a) 当 $n=3$ 时，列举所有的可能结果.

(b) 定义 $N(0)=1$，不用计算证明

$$N(n) = \sum_{i=1}^{n} \binom{n}{i} N(n-i)$$

提示：i 个选手和最后一名平局有多少种可能结果？

(c) 证明等式(b)和下列等式相同：

$$N(n) = \sum_{i=0}^{n-1} \binom{n}{i} N(i)$$

(d) 用递归方法求出 $N(3)$和 $N(4)$.

17. 对$\binom{n}{r}=\binom{n}{r,n-r}$给出组合解释.

18. 证明

$$\binom{n}{n_1,n_2,\cdots,n_r}=\binom{n-1}{n_1-1,n_2,\cdots,n_r}+\binom{n-1}{n_1,n_2-1,\cdots,n_r}+\cdots+\binom{n-1}{n_1,n_2,\cdots,n_r-1}$$

提示：与证明式(4-1)的方法类似.

19. 证明多项式定理.

*20. 有多少种方法把n个相同的球分到r个箱中，使得对每一个$i=1,\cdots,r$，第i个箱中至少有m_i个球？假设 $n \geqslant \sum_{i=1}^{r} m_i$.

21

*21. 证明 $x_1+x_2+\cdots+x_r=n$ 正好有$\binom{r}{k}\binom{n-1}{n-r+k}$个解，其中正好有 k 个 x_i 等于 0.

*22. 考虑含有 n 个变量的函数 $f(x_1,\cdots,x_n)$，它有多少个不同的 r 阶偏导数？

*23. 计算向量$(x_1,\cdots,x_n)$ 的个数，其中每一个 x_i 都是非负整数，且$\sum_{i=1}^{n} x_i \leqslant k$.

自测题与练习

1. 字母 A，B，C，D，E，F 有多少种不同的线性排列？

(a) A，B 相邻.

(b) A 在 B 前.

(c) A 在 B 前，且 B 在 C 前.

(d) A 在 B 前，且 C 在 D 前.

(e) A，B 相邻，且 C，D 相邻.

(f) E 不排在最后.

2. 4 个美国人、3 个法国人、3 个英国人坐成一排，如果相同国籍的人必须相邻而坐，有多少种坐法？

3. 从含 10 个人的俱乐部中选择一个主席、一个财务主管和一个秘书，且不兼职，有多少种选法？

(a) 没有限制.

(b) A，B 不能一起共事.

(c) C，D 要么一起选，要么都不选.

(d) E 必选.

(e) F 只有是主席时才能选择.

4. 在一次测试中，一个学生要回答 10 个问题中的 7 个问题，有多少种选法？如果前 5 道题必须至少回答 3 道，又有多少种选法？

5. 如果某人要将 7 个礼物分给他的 3 个孩子，如果最大的孩子分得 3 个礼物，其余两个孩子每人分得 2 个礼物，有多少种分法？

6. 如果登记的 7 位牌照中有 3 个字母、4 个数字，那么有多少种不同的牌照？假设允许字母和数字重复，并且它们的位置没有限制.

7. 给出下列等式的组合解释：

$$\binom{n}{r}=\binom{n}{n-r}$$

8. 考虑 n 位数的数，其中每一位数都取自 $0,1,\cdots,9$ 这 10 个数中的一个，有多少数满足下列条件？

(a) 没有两个相邻的数字是相同的.

(b) 0 出现 i 次，$i=0,\cdots,n$.

9. 考虑 3 个班级，每个班有 n 个学生，从这 $3n$ 个学生中抽取 3 个人.

(a) 有多少种选择方法？

(b) 选出的 3 个学生来自同一个班级，有多少种选法？

(c) 选出的 3 个学生中有 2 个来自一个班，另外 1 个来自不同的班，有多少种选法？

(d) 选出的 3 个学生来自不同的班，有多少种选法？

(e) 使用(a)到(d)的结果，给出一个组合等式.

22

10. $1,2,\cdots,9$ 可以组成多少个 5 位数？其中每一位数字最多出现 2 次(例如，41 434 就不满足条件).

11. $2n$ 个运动员参加一次乒乓球比赛，在第一轮比赛中，运动员被分成 n 对，每一对进行一场比赛，那么第一轮有多少种不同的结果？其中结果表明这 n 对产生了 n 个冠军.

12. 从 7 位男士、8 位女士中选出 6 人组成一个委员会，如果此委员会至少有 3 位女士和 2 位男士，可能产生多少种不同的委员会？

*13. 一个艺术收藏拍卖会要拍卖 4 幅达利的画、5 幅梵高的画、6 幅毕加索的画，并且拍卖的是 5 位收藏家收藏的作品. 如果一个记者注意到，每一个收藏家只有达利、梵高和毕加索的画的数目，那么当所有的作品都卖掉时有多少种不同的结果被报道？

*14. 求向量 $(x_1,\cdots,x_n)$ 的个数，使得每一个 x_i 都是正整数，且 $\sum_{i=1}^{n}x_i\leqslant k,k\geqslant n$.

*15. 共有 n 个学生参加概率中的保险精算考试的复习课，考试的结果只按成绩从高到低列出那些通过的学生，例如，如果只有 Brown 和 Cho 是通过考试的学生，并且 Brown 得分最高，那么结果就是“Brown，Cho”，假设所有的成绩均不相同(没有相同的分数)，那么有多少种可能的结果？

23

第2章　概率论的公理

2.1　引言

本章我们先介绍事件的概率的概念，然后说明在特定的条件下如何计算这些概率. 开始，有必要引入样本空间和试验的事件的概念.

2.2　样本空间与事件

考虑一个试验，其结果事先不能准确地预言. 然而，尽管试验的结果事先不知道，但我们假定所有可能结果的集合是已知的. 一个试验的所有可能结果的集合称为样本空间，记为 S. 下面举几个例子.

1. 若试验的结果是确定新生婴儿的性别，则

$$S=\{g,b\}$$

其中结果 g 表示婴儿为女孩，结果 b 表示婴儿为男孩.

2. 若一个试验的结果是 7 匹马的比赛结束时每匹马的名次，而马的编号分别为 1，2，3，4，5，6，7，则

$$S=\{7!\text{ 个}(1,2,3,4,5,6,7)\text{的全排列数}\}$$

例如，结果(2，3，1，6，5，4，7)表示 2 号马第一个到达，3 号马第二个到达，1 号马第三个到达，依此类推.

3. 考虑同时抛两枚硬币的试验，则样本空间由下列四个点组成：

$$S=\{(H,H),(H,T),(T,H),(T,T)\}$$

如果两枚硬币均为正面，则结果是(H,H)，(H,T)表示第一枚硬币为正面，第二枚硬币为反面，(T,H)表示第一枚硬币为反面，第二枚硬币为正面，(T,T)表示两枚硬币均为反面.

4. 若试验是同时抛两个骰子，则样本空间由 36 个点组成，即

$$S=\{(i,j):i,j=1,2,3,4,5,6\}$$

其中结果(i,j)表示第一个骰子出现 i 点，第二个骰子出现 j 点.

5. 若试验是测量某晶体管的寿命(单位：小时)，则样本空间由全体非负实数组成，即

$$S=\{x:0\leqslant x<\infty\}$$

样本空间的任意子集 E 称为事件，即一个事件是试验的可能结果组成的集合. 若试验的结果包含在 E 中，我们说 E 发生. 下面举几个事件的例子.

在前面的例 1 中，若 $E=\{g\}$，则 E 为婴儿是女孩的事件. 同样，若 $F=\{b\}$，则 F 为婴儿是男孩的事件.

在例 2 中，若

$$E=\{\text{在 } S \text{ 中以 3 开始的所有结果}\}$$

则 E 表示在比赛中 3 号马获胜的事件.

在例 3 中，若 $E=\{(H,H),(H,T)\}$，则 E 表示第一枚硬币出现正面的事件.

在例 4 中，若 $E=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$，则 E 表示两个骰子的点数之和等于 7 的事件.

在例 5 中，若 $E=\{x:0\leqslant x\leqslant 5\}$，则 E 表示此晶体管的寿命不超过 5 小时的事件.

对样本空间 S 的任意两个事件 E 和 F，我们定义新事件 $E\cup F$，由在 E 中、在 F 中或者在 E 与 F 之中的所有结果组成. 即事件 $E\cup F$ 出现，如果 E 出现或 F 出现. 例如，在例 1 中，若事件 $E=\{g\}$和 $F=\{b\}$，则

$$E\cup F=\{g,b\}$$

即 $E\cup F$ 是整个样本空间 S. 在例 3 中，若 $E=\{(H,H),(H,T)\}$和 $F=\{(T,H)\}$，则

$$E\cup F=\{(H,H),(H,T),(T,H)\}$$

因此，如果任何一枚硬币出现正面，则 $E\cup F$ 发生.

事件 $E\cup F$ 称为事件 E 与事件 F 的**并**.

类似地，对任意两个事件 E 和 F，我们可以定义新事件 EF，称为事件 E 和 F 的**交**，它由既在 E 中又在 F 中的所有结果组成，即事件 EF(有时也写作 $E\cap F$)发生当且仅当 E 与 F 同时发生. 例如，在例 3 中，若 $E=\{(H,H),(H,T),(T,H)\}$是至少一枚硬币出现正面的事件，$F=\{(H,T),(T,H),(T,T)\}$是至少一枚硬币出现反面的事件，则

25

$$EF=\{(H,T),(T,H)\}$$

是恰有一枚硬币出现正面和一枚硬币出现反面的事件. 在例 4 中，若 $E=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$是两个骰子的点数之和为 7 的事件，$F=\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$是两个骰子的点数之和为 6 的事件，则 EF 不包含任何结果，因此事件不发生. 为给这个事件起一个名字，我们称其为不可能事件，记为$\varnothing$(即$\varnothing$表示不包含任何结果的事件). 如果 $EF=\varnothing$，则称 E 和 F 为**互不相容的**.

类似地，我们可以定义两个以上事件的并与交. 若 $E_1,E_2,\cdots$ 是一系列事件，这些事件的并 $\bigcup\limits_{n=1}^{\infty} E_n$ 定义为由所有至少属于某个$E_n(n=1,2,\cdots)$的结果组成的事件. 类似地，这些事件的交 $\bigcap\limits_{n=1}^{\infty} E_n$ 定义为由属于所有 $E_n(n=1,2,\cdots)$ 的结果组成的事件.

最后，对任意事件 E 我们定义一个新事件 E^c，称为事件 E 的**余事件**，表示在样本空间 S 中但不在 E 中的所有结果组成的事件，即 E^c 发生当且仅当 E 不发生. 在例 4 中，若事件 $E=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$，当两个骰子的点数之和不等于 7 时，则 E^c 发生. 注意到试验必须得到某些结果，由此推出 $S^c=\varnothing$.

对任意两个事件 E 和 F，若 E 中的所有结果也在 F 中，则称 E 包含在 F 中，记作 $E\subset F$(或等价地，$F\supset E$). 因此，若 $E\subset F$，则 E 的出现必然蕴涵着 F 的出现. 若 $E\subset F$ 且 $F\subset E$，则称 E 和 F 是相等的，记作 $E=F$.

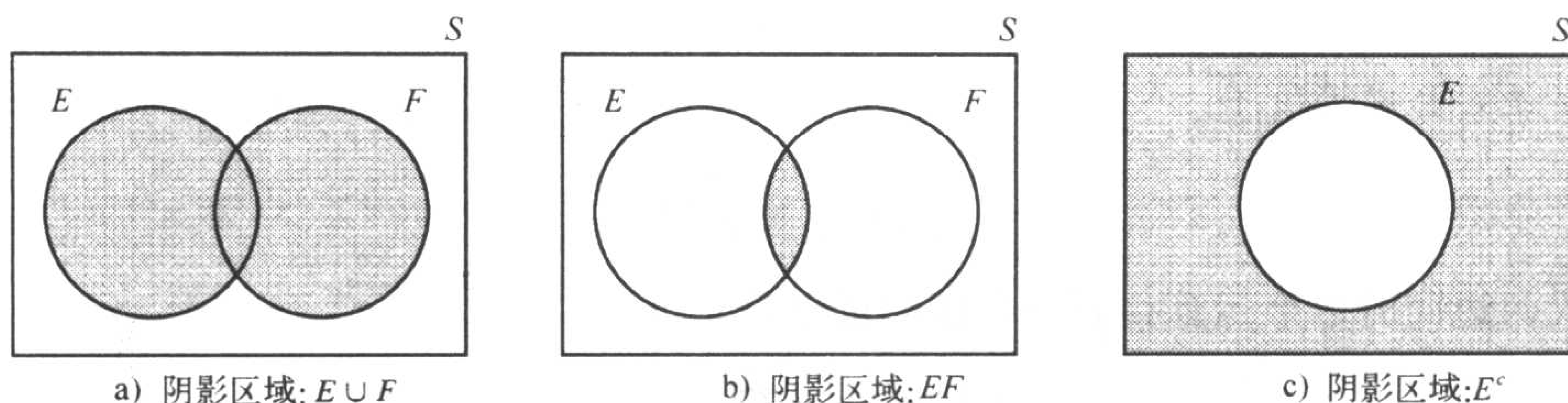
文氏图可用来很好地说明事件之间的逻辑关系. 用一个大的矩形内的所有结果来表示样本空间 S，用矩形内给定的圆内的结果分别表示事件 $E,F,G,\cdots$，图中的阴影区域代表感兴趣的事件. 例如图 2-1 中的三个文氏图，阴影部分分别表示事件 $E\cup F$、EF 和 E^c，图 2-2 中的文氏图表示 $E\subset F$.

事件的并、交、余形成的运算类似于代数的运算，服从一定的运算规律，下面列出几个规律：

交换律　$E\cup F=F\cup E$　　$EF=FE$

结合律　$(E\cup F)\cup G=E\cup(F\cup G)$　　$(EF)G=E(FG)$

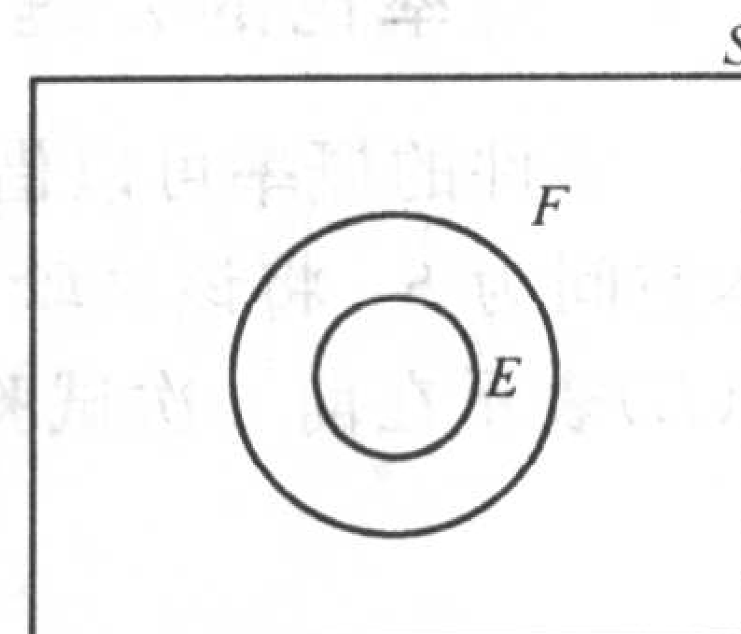
分配律　$(E\cup F)G=EG\cup FG$　　　$EF\cup G=(E\cup G)(F\cup G)$



a) 阴影区域: $E\cup F$　b) 阴影区域: EF　c) 阴影区域: E^c

图 2-1

要验证这些关系式，只要证明包含在等式左边的事件的任意结果也包含在等式右边的事件中，反之也成立. 另一种证明的方法是利用文氏图. 例如，利用图 2-3 中的一系列示意图可以验证分配律.



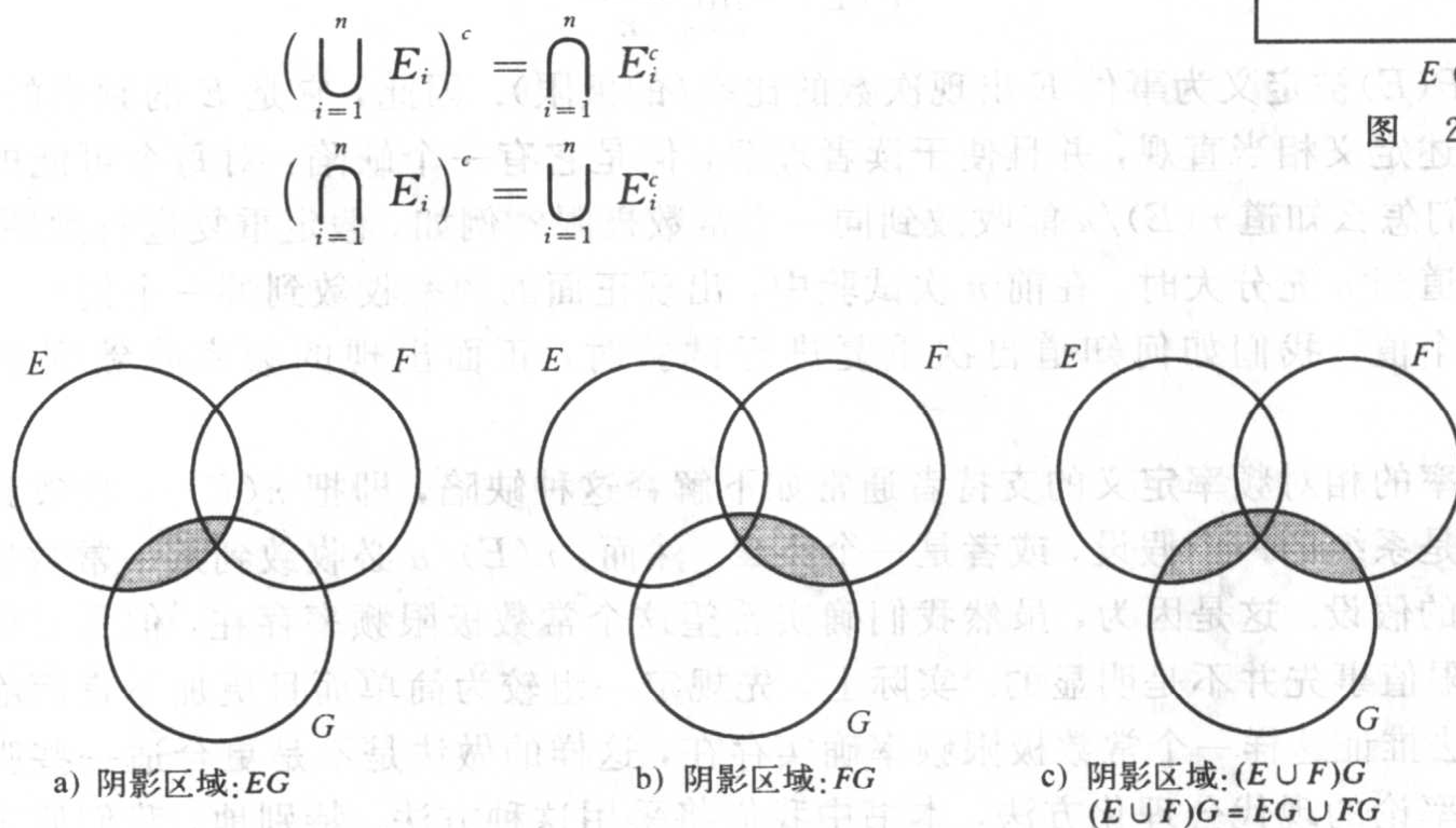
$E\subset F$

图 2-2

下列并、交、余三种基本运算之间最有用的关系式是所谓的德摩根律：

$$\left(\bigcup_{i=1}^{n} E_i\right)^c=\bigcap_{i=1}^{n} E_i^c$$

$$\left(\bigcap_{i=1}^{n} E_i\right)^c=\bigcup_{i=1}^{n} E_i^c$$



a) 阴影区域: EG　b) 阴影区域: FG　c) 阴影区域: $(E\cup F)G$

$(E\cup F)G=EG\cup FG$

图 2-3

为了证明德摩根律，首先假定 x 是事件 $\left(\bigcup_{i=1}^{n} E_i\right)^c$ 的任一结果，则 x 不包含在 $\bigcup_{i=1}^{n} E_i$ 中，这就意味着 x 不包含在事件 $E_i(i=1,2,\cdots)$ 中，由此推出对一切 $i=1,2,\cdots,n$，x 包含在 E_i^c 中，因此 x 包含在 $\bigcap_{i=1}^{n} E_i^c$ 中. 相反，假定 x 是事件 $\bigcap_{i=1}^{n} E_i^c$ 的一个结果，则对一切 $i=1,2,\cdots,n$，x 包含在 E_i^c 中，这就意味着对一切 $i=1,2,\cdots,n$，x 不包含在 E_i 中，由此推出 x 不包含在 $\bigcup_{i=1}^{n} E_i$ 中，因此 x 包含在 $\left(\bigcup_{i=1}^{n} E_i\right)^c$ 中. 这就证明了德摩根律的第一个等式.

为了证明德摩根律的第二个等式，我们利用第一个等式得到

$$\left(\bigcup_{i=1}^{n} E_i^c\right)^c = \bigcap_{i=1}^{n} (E_i^c)^c$$

因为$(E^c)^c = E$，上式等价于

$$\left(\bigcup_{i=1}^{n} E_i^c\right)^c = \bigcap_{i=1}^{n} E_i$$

对上面等式两边同时取余运算得到所要证明的结论，即

$$\bigcup_{i=1}^{n} E_i^c = \left(\bigcap_{i=1}^{n} E_i\right)^c$$

2.3 概率论的公理

26 ~ 28

事件的概率可以借助于它的相对频率来定义. 这种定义通常叙述为：设有一个试验，其样本空间为 S，将该试验在完全相同的条件下重复进行. 对样本空间 S 中的任一个事件 E，用 $n(E)$表示在前 n 次试验中事件 E 出现的次数，那么事件 E 的概率 $P(E)$定义为：

$$P(E) = \lim_{n\to\infty}\frac{n(E)}{n}$$

这就是说，$P(E)$被定义为事件 E 出现次数的比率(的极限). 因此，它是 E 的频率的极限.

尽管上述定义相当直观，并且便于读者理解，但是它有一个缺陷：对每个可能的试验的重复序列，我们怎么知道 $n(E)/n$ 能收敛到同一个常数极限？例如，假定重复进行抛硬币的试验，我们怎么知道当 n 充分大时，在前 n 次试验中，出现正面的频率收敛到哪一个值？并且即使它收敛到某一个值，我们如何知道再次重复进行试验时，正面出现的频率收敛到相同的极限值呢？

有关概率的相对频率定义的支持者通常如下解释这种缺陷，即把 $n(E)/n$ 收敛到某一常数极限值看成是系统的一个假设，或者是一个公理. 然而，$n(E)/n$ 必收敛到某一常数极限值是一个十分复杂的假设. 这是因为，虽然我们确实希望这个常数极限频率存在，但是对所需要的情形，这个极限值事先并不是明显的. 实际上，先规定一组较为简单而且更加不言而喻的概率公理，然后设法推证这样一个常数极限频率确实存在，这样的做法是不是更合适一些呢？后一种方式正是概率论的现代公理化方法，本书中我们将采用这种方法. 特别地，我们假定对样本空间 S 中的每一个事件 E，存在一个数值 $P(E)$，称为事件 E 的概率. 我们将假定概率满足一组特定的公理，这些公理与概率的直观定义是一致的，关于这一点希望得到读者的赞同.

考虑样本空间为 S 的一个试验，对于样本空间 S 的每一个事件 E，假设定义了量 $P(E)$且满足下面三个公理.

公理 1

$$0 \leqslant P(E) \leqslant 1$$

公理 2

29

$$P(S) = 1$$

公理 3 对任意两两互不相容的事件序列 $E_1, E_2, \cdots$(即 $E_iE_j = \varnothing$，$i \neq j$)，有

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

我们称 $P(E)$为事件 E 的概率.

因此，公理 1 说明，在 E 中试验结果的概率是 0 与 1 之间的某一个数. 公理 2 说明，全部试验结果是样本空间 S 中的点的概率为 1. 公理 3 说明，两两互不相容的事件序列至少有一个事件发生的概率正好等于它们各自的概率的和.

假如我们考虑一个事件序列 $E_1,E_2,\cdots$，其中 $E_1=S$，$E_i=\varnothing$，$i>1$，那么这些事件是两两互不相容的，且 $S=\bigcup_{i=1}^{\infty}E_i$，由公理 3 可以得到

$$P(S)=\sum_{i=1}^{\infty}P(E_i)=P(S)+\sum_{i=2}^{\infty}P(\varnothing)$$

由此推出

$$P(\varnothing)=0$$

即不可能事件出现的概率等于 0.

对任意有限个两两互不相容的事件序列 $E_1,E_2,\cdots,E_n$，可以得到

$$P\left(\bigcup_{i=1}^{n}E_i\right)=\sum_{i=1}^{n}P(E_i) \tag{3-1}$$

由公理 3，对一切 $i>n$，只要取 $E_i=\varnothing$即可. 当样本空间有限时，公理 3 等价于式(3-1).（为什么?）当样本空间由无限个点组成时，公理 3 的可加性是必要的.

例 3a 假定试验是抛一枚硬币，且假设硬币是均匀的，则有

$$P(\{H\})=P(\{T\})=\frac{1}{2}$$

另一方面，假定硬币是偏重的，且我们认为正面出现的机会是反面的两倍，则有

$$P(\{H\})=\frac{2}{3}\qquad P(\{T\})=\frac{1}{3}$$ ■

30

例 3b 假如掷一个骰子，且假定 6 个面出现的机会相等，则有 $P(\{1\})=P(\{2\})=P(\{3\})=P(\{4\})=P(\{5\})=P(\{6\})=\frac{1}{6}$，因此由公理 3 可得掷出偶数点的概率等于

$$P(\{2,4,6\})=P(\{2\})+P(\{4\})+P(\{6\})=\frac{1}{2}$$ ■

定义在样本空间 S 上且满足公理 1，2 和 3 的集函数 P 的存在性的假设构成了概率论的现代数学方法. 希望读者感到引入公理化定义是自然的，并且与机会和随机性相关的概率的直观概念相吻合. 而且，利用公理可以证明：如果试验能一次又一次地重复进行，那么，任意特定的事件 E 出现的比率以概率 1 将是 $P(E)$. 这个结果称为强大数定律，将在第 8 章介绍. 另外，我们将在 2.7 节给出概率的另一种解释——作为置信的度量.

注 我们已经假定 $P(E)$定义在样本空间的所有事件上. 事实上，当样本空间是不可数无限集时，$P(E)$仅仅定义在一类所谓可测事件集上. 然而，我们并不关心这个限制，因为任何我们特别感兴趣的事件都是可测的.

2.4 一些简单命题

本节我们证明一些关于概率的简单命题. 首先注意到 E 和 E^c 是互不相容的事件，且 $E\cup E^c=S$，由公理 2 和公理 3 得

$$1=P(S)=P(E\cup E^c)=P(E)+P(E^c)$$

或者，等价地叙述为如下的命题 4.1.

命题 4.1

$$P(E^c)=1-P(E)$$

总之，命题 4.1 说明一个事件不出现的概率等于 1 减去该事件出现的概率. 例如，假如抛一枚硬币出现正面的概率是$\frac{3}{8}$，则出现反面的概率等于$\frac{5}{8}$.

第二个命题是说，若事件 E 包含在事件 F 之中，则事件 E 的概率不大于事件 F 的概率.

31

命题 4.2 若 $E\subset F$，则 $P(E)\leqslant P(F)$.

证明 因为 $E\subset F$，所以 F 可以表示为

$$F=E\cup E^cF$$

而 E 和 E^cF 互不相容，由公理 3 可得

$$P(F)=P(E)+P(E^cF)$$

因为 $P(E^cF)\geqslant 0$，命题得证. ■

命题 4.2 告诉我们，比如说，掷一个骰子出现 1 点的概率小于等于出现奇数点的概率.

下一个命题给出两个事件的并的概率与它们各自的概率以及它们的交的概率之间的关系.

命题 4.3

$$P(E\cup F)=P(E)+P(F)-P(EF)$$

证明 为了得到 $P(E\cup F)$的公式，首先注意到 $E\cup F$ 可以表示为两个不相交事件 E 和 E^cF 的并，因此由公理 3 得到

$$P(E\cup F)=P(E\cup E^cF)=P(E)+P(E^cF)$$

另一方面，因为 $F=EF\cup E^cF$，从公理 3 又得到

$$P(F)=P(EF)+P(E^cF)$$

或者等价地，

$$P(E^cF)=P(F)-P(EF)$$

于是命题得证.

还可以利用图 2-4 中的文氏图来证明命题 4.3.

如图 2-5 所示，我们将 $E\cup F$ 分为三个互不相容的部分，即部分Ⅰ表示在 E 中而不在 F 中的所有点(即 EF^c)，部分Ⅱ表示既在 E 中又在 F 中的所有点(即 EF)，部分Ⅲ表示在 F 中而不在 E 中的所有点(即 E^cF).

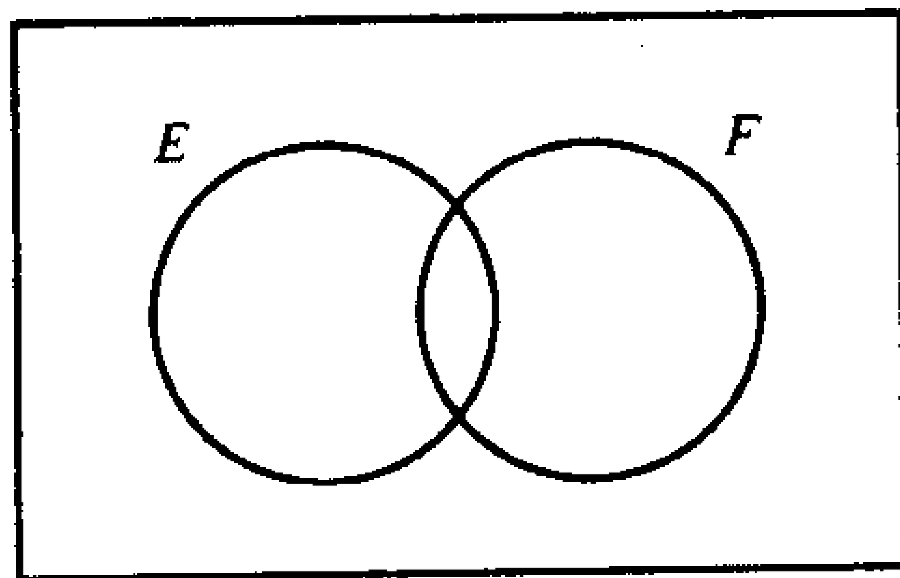


图 2-4 文氏图

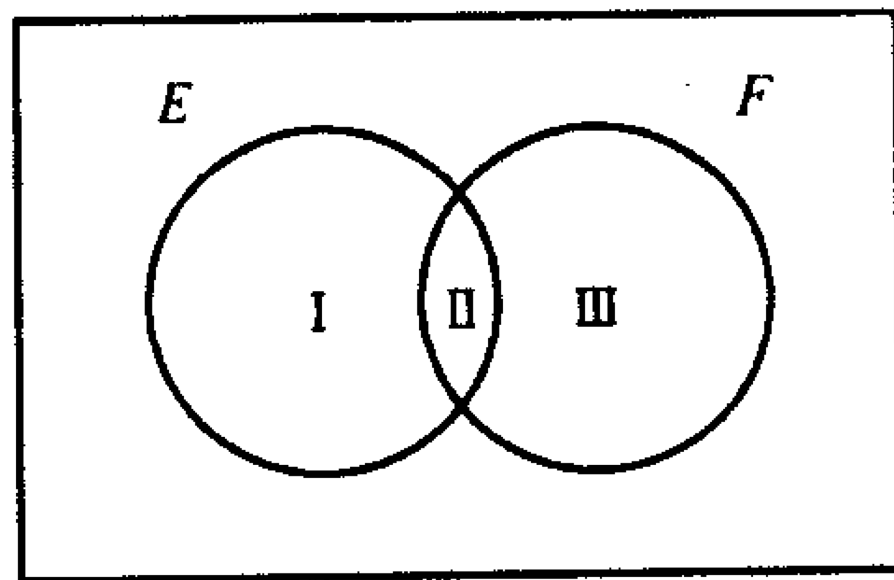


图 2-5 分块文氏图

从图 2-5 可以看出

$$E\cup F=\text{Ⅰ}\cup\text{Ⅱ}\cup\text{Ⅲ}$$

$$E = \text{I} \cup \text{II}$$
$$F = \text{II} \cup \text{III}$$

因为Ⅰ，Ⅱ，Ⅲ是互不相容的，由公理 3 得到

$$P(E \cup F) = P(\text{I}) + P(\text{II}) + P(\text{III})$$
$$P(E) = P(\text{I}) + P(\text{II})$$
$$P(F) = P(\text{II}) + P(\text{III})$$

这就证明了

$$P(E \cup F) = P(E) + P(F) - P(\text{II})$$

因为Ⅱ $= EF$，命题 4.3 得证. ■

例 4a 若抛两枚硬币，且假定样本空间 $S = \{(H,H),(H,T),(T,H),(T,T)\}$ 中的 4 个点是等可能的，因此概率是 $\frac{1}{4}$. 设

$$E = \{(H,H),(H,T)\} \qquad F = \{(H,H),(T,H)\}$$

即 E 是第一枚硬币出现正面的事件，F 是第二枚硬币出现正面的事件.

由命题 4.3，第一枚硬币或第二枚硬币出现正面的概率 $P(E \cup F)$ 为

$$P(E \cup F) = P(E) + P(F) - P(EF) = \frac{1}{2} + \frac{1}{2} - P(\{(H,H)\}) = 1 - \frac{1}{4} = \frac{3}{4}$$

当然，这个概率可以直接计算，即

$$P(E \cup F) = P(\{(H,H),(H,T),(T,H)\}) = \frac{3}{4}$$ ■

32 ~ 33

还可以计算三个事件 E、F 或 G 中的任一个出现的概率：

$$P(E \cup F \cup G) = P[(E \cup F) \cup G]$$

由命题 4.3，这个概率等于

$$P(E \cup F) + P(G) - P[(E \cup F)G]$$

由分配律可知事件 $(E \cup F)G$ 和事件 $EG \cup FG$ 是相等的，因此由前面的等式得到

$$\begin{aligned} P(E \cup F \cup G) &= P(E) + P(F) - P(EF) + P(G) - P(EG \cup FG) \\ &= P(E) + P(F) - P(EF) + P(G) - P(EG) - P(FG) + P(EGFG) \\ &= P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG) \end{aligned}$$

事实上，由归纳法可以得到以下命题.

命题 4.4

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \cdots + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} E_{i_2} \cdots E_{i_r}) + \cdots + (-1)^{n-1} P(E_1 E_2 \cdots E_n)$$

其中，$\sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} E_{i_2} \cdots E_{i_r})$ 表示对集合 $\{1, 2, \cdots, n\}$ 的所有 $\binom{n}{r}$ 种大小为 r 的可能的子集求和.

命题 4.4 说明，n 个事件并的概率等于一次取一个事件的这些事件概率的和，减去一次取两个事件的这些事件概率的和，再加上一次取三个事件的这些事件概率的和，以此类推.

注 为得到命题 4.4 的非诱导结论，首先注意假如样本空间的结果不是任意一个集 E_i 的成员，则其概率不对等式两边贡献任何东西. 另一方面，假定一个结果正好在 m 个

事件 E_i 中，其中 $m>0$，那么，该结果在 $\bigcup_i E_i$ 中，其概率在 $P\left(\bigcup_i E_i\right)$ 中一次一次计算；还因为这个结果包含在类型为 $E_{i_1}E_{i_2}\cdots E_{i_k}$ 的 $\binom{m}{k}$ 个子集中，所以其概率在命题

34

4.4 中的等式右边被计算了

$$\binom{m}{1}-\binom{m}{2}+\binom{m}{3}-\cdots\pm\binom{m}{m}$$

次. 于是，对 $m>0$，我们必须证明

$$1=\binom{m}{1}-\binom{m}{2}+\binom{m}{3}-\cdots\pm\binom{m}{m}$$

然而，因为 $1=\binom{m}{0}$，所以上面的等式等价于

$$\sum_{i=0}^{m}\binom{m}{i}(-1)^i=0$$

后一个方程服从二项式定理，因为

$$0=(-1+1)^m=\sum_{i=0}^{m}\binom{m}{i}(-1)^i(1)^{m-i}$$

2.5 具有等可能结果的样本空间

对许多试验，假定样本空间的所有结果等可能出现是很自然的，即考虑一个试验，其样本空间是有限集，比如说 $S=\{1,2,\cdots,N\}$，那么，很自然地假定

$$P(\{1\})=P(\{2\})=\cdots=P(\{N\})$$

由公理 2 和公理 3 得到(为什么?)

$$P(\{i\})=\frac{1}{N}\qquad i=1,2,\cdots,N$$

再次用公理 3 可知，对任意事件 E，

$$P(E)=\frac{E\text{ 中点的个数}}{S\text{ 中点的个数}}$$

简言之，如果假定一个试验的所有结果是等可能出现的，则任意事件 E 的概率等于 E 所包含的试验结果的个数与样本空间的试验结果的个数之比.

例 5a 掷两个骰子，它们出现的点数之和为 7 的概率是多少?

解 解此题时，我们假定 36 个可能结果中的每一个是等可能的. 因为点数之和等于 7 含

35

有 6 个可能结果(1, 6)，(2, 5)，(3, 4)，(4, 3)，(5, 2)，(6, 1)，故所求概率为 $\frac{6}{36}=\frac{1}{6}$. ■

例 5b 从一只装有 6 个白球和 5 个黑球的碗中随机地抽取 3 个球，问所抽取的球中正好一个为白球而另外两个为黑球的概率是多少?

解 若认为球的抽取与次序是有关的，则样本空间包含了 $11\cdot10\cdot9=990$ 个结果. 所抽取的第一个球为白球而其余两球均为黑球有 $6\cdot5\cdot4=120$ 个结果；第一个球为黑球、第二个球为白球而第三个球为黑球有 $5\cdot6\cdot4=120$ 个结果；前两球为黑球而第三个球为白球有 $5\cdot4\cdot6=120$ 个结果. 假定“随机抽取”表示每个结果在样本空间中是等可能出现的，则所求概率为

$$\frac{120+120+120}{990}=\frac{4}{11}$$

我们在抽球的集合是无序的情形下来解此题. 从这个意义上讲，样本空间中包含了$\binom{11}{3}=165$个结果. 现在，当考虑选择的次序时，3 个球的每一个集合对应于 3! 个结果. 因此，如果考虑选择的次序，假定所有的结果是等可能出现的，那么只有当结果是从球的无序集合中抽取时等可能性才保持. 因此，所求概率为

$$\frac{\binom{6}{1}\binom{5}{2}}{\binom{11}{3}}=\frac{4}{11}$$

当然，这与前面的结果是一致的. ■

例 5c 从 6 个男人和 9 个女人的小组中选出 5 个人组成一个委员会，假定选取是随机的，问委员会正好由 3 个男人和 2 个女人组成的概率是多少?

解 假定随机选取就意味着$\binom{15}{5}$种可能组合中的每一种都等可能被选到. 因此，所求概率为

$$\frac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}}=\frac{240}{1\,001}$$

■ 36

例 5d 一个罐中装有 n 个球，其中只有一个球有标记. 若一次从罐中抽出 k 个球，罐中的每一个球同时被选到是等可能的，有标记的球被选到的概率是多少?

解 因为以同等的方式来对待所有的球，由此推出所选择的 k 个球的集合是$\binom{n}{k}$个集合中的任意一个集合，因此

$$P\{\text{选到有标记的球}\}=\frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}}=\frac{k}{n}$$

通过令 A_i 表示有标记的球为第 i 次所选择的球的事件，$i=1,2,\cdots,k$，我们也可得到前面的结果. 因为 n 个球中的每一个球都等可能的是第 i 个被选择的球，因此 $P(A_i)=\frac{1}{n}$. 显然这些事件是互不相容的，因此我们得到

$$P\{\text{选到有标记的球}\} = P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k}P(A_i) = \frac{k}{n}$$

$P(A_i)=1/n$ 也可用以下方法获得，注意到总共有 $n(n-1)\cdots(n-k+1)=n!\,/(n-k)!$ 个等可能的试验结果，有标记的球是第 i 次所选择的球的事件有$(n-1)(n-2)\cdots(n-i+1)(1)(n-i)\cdots(n-k+1)=(n-1)!\,/(n-k)!$ 个结果，由此得出

$$P(A_i)=\frac{(n-1)!}{n!}=\frac{1}{n}$$

■

例 5e 假定将 $n+m$ 个球(其中有 n 个红球、m 个蓝球)以线性次序排列，所有 $(n+m)!$ 种排列方式是等可能的. 若我们只记录球的颜色相间的试验结果，证明所有可能结果保持等可能性.

解 考虑 $(n+m)!$ 种可能排列的任意一种排列，注意到其中红球的任意一种排列和蓝球的任意一种排列不改变颜色的次序. 因此，颜色的每一种排序对应 $(m+n)$ 个球的 $n!\ m!$ 种不同排序. 所以，每种颜色依次出现的概率为 $\frac{n!\ m!}{(n+m)!}$.

例如，假定有 2 个红球(记为 r_1, r_2)和 2 个蓝球(记为 b_1, b_2)，那么，在 4! 种可能的排列中，任一种特定颜色的组合将有 2! 2! 种排列. 例如，下面的排列是球的颜色交替而红球首先出现时的排列结果：

$$r_1, b_1, r_2, b_2 \qquad r_1, b_2, r_2, b_1 \qquad r_2, b_1, r_1, b_2 \qquad r_2, b_2, r_1, b_1$$

37

因此颜色的每一种可能的排列出现的概率为 $\frac{4}{24}=\frac{1}{6}$. ■

例 5f 一手扑克牌由 5 张牌组成，如果这手牌由 5 个连贯的点数和不是同一花色组成就称为"顺子". 例如，由黑桃 5、黑桃 6、黑桃 7、黑桃 8 和红心 9 组成的一手牌就是一个"顺子". "顺子"出现的概率是多少?

解 假定所有 $\binom{52}{5}$ 手牌是等可能的. 为了计算所有"顺子"的总数，我们首先计算一手牌由 A，2，3，4 和 5(不管花色)组成共有多少种可能结果. 因为 A 可以是 4 个 A 中的任何一个，2，3，4 和 5 也一样，由此得到正好由一个 A，2，3，4 和 5 组成的一手牌共有 4^5 种可能结果. 因为在所有手牌的结果中有 4 种相同的花色(这手牌叫做"同花顺")，由此推出有 4^5-4 种"顺子"是由 A，2，3，4 和 5 组成. 同理也有 4^5-4 种"顺子"是由 10，J，Q，K 和 A 组成. 因此共有 $10(4^5-4)$ 种"顺子". 于是所求概率为

$$\frac{10(4^5-4)}{\binom{52}{5}} \approx 0.003\,9$$ ■

例 5g 对于 5 张扑克牌的一手牌，如果其中的 3 张同点和另外 2 张同点，我们称它为"满堂红"(即一个"满堂红"由同一点数的 3 张牌和一对牌组成). "满堂红"出现的概率是多少?

解 假定所有 $\binom{52}{5}$ 手牌是等可能的. 为了确定所有"满堂红"的总数，首先注意到有 $\binom{4}{2}\binom{4}{3}$ 种不同的组合，比如说 2 个 10 和 3 个 J. 因为一对牌的选法有 13 种，一对牌取定后，其余 3 张同点数的牌的选法有 12 种，由此得到"满堂红"出现的概率为

$$\frac{13 \cdot 12 \cdot \binom{4}{2} \cdot \binom{4}{3}}{\binom{52}{5}} \approx 0.001\,4$$ ■

例 5h 在桥牌游戏中将全部 52 张纸牌分发给 4 个人，求下列事件的概率：

(a) 一人得到全部 13 张黑桃.

38

(b) 每个人得到一张 A.

解　(a) 对 4 个人，纸牌共有$\binom{52}{13,13,13,13}$种可能的分法，某指定的人得到全部 13 张黑桃纸牌有$\binom{39}{13,13,13}$种可能的分法，故所求概率为

$$\frac{4\binom{39}{13,13,13}}{\binom{52}{13,13,13,13}}\approx 6.3\times 10^{-12}$$

(b) 为了确定每个人均得到一个 A 的结果数，把 4 个 A 放在一边，将其余 48 张牌分给 4 个人，每人 12 张，共有$\binom{48}{12,12,12,12}$种不同的分法. 因为每人得到一个 A 共有 4! 种分法，所以每个人恰好有一个 A 的可能结果数为 $4!\binom{48}{12,12,12,12}$. 因此所求概率为

$$\frac{4!\binom{48}{12,12,12,12}}{\binom{52}{13,13,13,13}}\approx 0.105$$

■

在概率论中，有些结果在我们初次遇到它时是相当出乎预料的. 下面我们通过两个例子说明这种现象.

例 5i　若 n 个人分在一个房间，没有两个人在一年的同一天过生日的概率是多少？多大的 n 可使得这个概率小于$\frac{1}{2}$？

解　因为每个人在一年的 365 天中的任意一天过生日，所以共有$(365)^n$ 个可能结果(我们不考虑某人在 2 月 29 日出生的情况). 假定每一个结果是等可能的，可得到所求概率为$(365)(364)(363)\cdots(365-n+1)/(365)^n$. 人们惊奇地发现，当 $n\geqslant 23$ 时这个概率就小于$\frac{1}{2}$. 也就是说，若一个房间中的人数多于 23 人，那么至少有两个人的生日相同的概率大于$\frac{1}{2}$. 许多人最初对这个结果感到很奇怪，因为相对于一年 365 天来说 23 似乎太小了. 然而，每两个人中生日相同的概率是$\frac{365}{(365)^2}=\frac{1}{365}$，而 23 个人中共有$\binom{23}{2}=253$ 个不同的对. 注意到这个结果就不会再感到奇怪了.

39

当 50 个人在一个房间时，至少两个人生日相同的概率近似为 0.970，而当一个房间有 100 个人时，至少有两个人生日相同的几率大于 3 000 000 : 1$\left(\text{即概率大于}\frac{3\times 10^6}{3\times 10^6+1}\right)$.　■

例 5j　将 52 张扑克牌洗后一次翻一张牌，直到第一个 A 出现，下一张牌(即接着第一个 A 出现的牌)是黑桃 A 或者梅花 2 的概率是多少？

解　为了计算第一个 A 出现后接着出现黑桃 A 的概率，我们需要计算(52)! 种牌的可能排列中黑桃 A 紧接着第一个 A 出现的排列有多少种. 首先注意到 52 张牌的每一种排列可由将黑桃 A 插入与黑桃 A 不同的 51 张牌的一个排列中而得到. 而其余牌的(51)! 种不同的排列

中，只有一个位置是黑桃 A 紧接着第一个出现的 A 来排. 例如，若 51 张牌的排列为

梅花 4，红桃 6，方块 J，黑桃 5，梅花 A，方块 7，…，红桃 K

那么将黑桃 A 插入到以上排列中使其紧接着第一个 A 的排列是

梅花 4，红桃 6，方块 J，黑桃 5，梅花 A，黑桃 A，方块 7，…，红桃 K

因此，使得黑桃 A 接着第一个 A 出现的可能排列有(51)! 种，故

$$P\{\text{黑桃 A 接着第一个 A 出现}\}=\frac{(51)!}{(52)!}=\frac{1}{52}$$

事实上，同理可以推出梅花 2(或任何其他一张指定的牌)紧接着第一个 A 出现的概率也是 $\frac{1}{52}$. 换句话说，一副扑克牌的 52 张牌中每一张牌紧接着第一个 A 出现是等可能的.

许多人对这个结果感到十分惊讶. 的确一个共同的反应是最初推测梅花 2(而不是黑桃 A)紧接着第一个 A 出现的可能性大，因为第一个 A 本身可以是黑桃 A. 这种反应通常是认为梅花 2 本身可以在第一个 A 之前出现. 然而，黑桃 A 以第一个 A 出现的机会为四分之一(因为四个 A 作为第一个 A 出现是等可能的)，而梅花 2 将在第一个 A 之前出现的机会为五分之一(因为由梅花 2 和 4 个 A 组成的 5 张牌的集合中每一个首次出现是等可能的)，因此梅花 2 再一次出现的可能性大. 但是，这不是事实，通过更复杂的分析可以证明它们是等可能的. ■

40

例 5k 某足球队由攻击型队员和防守型队员各 20 名组成. 为确定哪两个人同住一屋，将队员配成两人一对，如果配对是随机的，那么攻击型队员与防守型队员不同住一屋的概率是多少? $2i(i=1,2,\cdots,10)$对攻击型队员与防守型队员同住一屋的概率是多少?

解 将 40 名队员每两人一对分成 20 个有序对(即把他们分为第一对，第二对，等等)，共有

$$\binom{40}{2,2,\cdots,2}=\frac{(40)!}{(2!)^{20}}$$

种分法. 因此，若不考虑对与对之间的次序(无需)，则有$\frac{(40)!}{2^{20}(20)!}$种分法. 此外，为了不使攻击型队员与防守型队员同住一屋，必须在攻击型队员之间、防守型队员之间各自分别配对，而这时有$\left[\frac{(20)!}{2^{10}(10)!}\right]^2$ 种配法. 因此，攻击型队员与防守型队员不同住一屋的概率(记为 P_0)为

$$P_0=\frac{\left[\frac{(20)!}{2^{10}(10)!}\right]^2}{\frac{(40)!}{2^{20}(20)!}}=\frac{[(20)!]^3}{[(10)!]^2(40)!}$$

为了计算 $2i$ 对攻击型队员与防守型队员同住一屋的概率 P_{2i}，我们首先指出，选择 $2i$ 个攻击型队员与 $2i$ 个防守型队员组成攻击型－防守型对的选法有$\binom{20}{2i}^2$ 种，然后将这 $4i$ 个队员配成攻击型－防守型对有$(2i)!$ 配法. (这是因为第一个攻击型队员可以与 $2i$ 个防守型队员中的任何一个配对，第二个攻击型队员可与余下的 $2i-1$ 个防守型队员中的任何一个配对，等等.)由于所剩的 $20-2i$ 个攻击型队员与 $20-2i$ 个防守型队员只能分别各自配对，所以恰有 $2i$ 对攻击型队员与防守型队员同住一屋的分法有

$$\binom{20}{2i}^2(2i)!\ \left[\frac{(20-2i)!}{2^{10-i}(10-i)!}\right]^2$$

种，于是得到

$$P_{2i}=\frac{\binom{20}{2i}^2(2i)!\left[\frac{(20-2i)!}{2^{10-i}(10-i)!}\right]^2}{\frac{(40)!}{2^{20}(20)!}}\qquad i=0,1,2,\cdots,10$$

41

现在我们可以计算或利用斯特林(Stirling)公式近似计算 $P_{2i}(i=0,1,2,\cdots,10)$，斯特林公式表明 $n!$ 近似等于 $n^{n+\frac{1}{2}}+\mathrm{e}^{-n}\sqrt{2\pi}$. 例如，我们得到

$$P_0\approx 1.340\,3\times 10^{-6}$$
$$P_{10}\approx 0.345\,861$$
$$P_{20}\approx 7.606\,8\times 10^{-6}$$

■

下面 3 个例子说明了命题 4.4 的应用. 在例 5l 中，概率的引入使我们很快得到计数问题的解.

例 5l 一个俱乐部的 36 个成员打网球，28 个成员打壁球，18 个成员打羽毛球. 22 个成员既打网球又打壁球，12 个成员既打网球又打羽毛球，9 个成员既打壁球又打羽毛球，4 个成员所有 3 种球都打. 问这个俱乐部中有多少个成员至少参加这 3 种运动中的一种?

解 用 N 表示该俱乐部中的成员总数，且假定俱乐部的成员随机地被选取. 若对俱乐部的任意一些成员的子集 C，我们用 $P(C)$ 表示包含在 C 中的成员被选取的概率，则

$$P(C)=\frac{C\text{ 中的成员数}}{N}$$

现在用 T 表示打网球的成员的集合，S 表示打壁球的成员的集合，B 表示打羽毛球的成员的集合，由命题 4.4 得

$$\begin{aligned}P(T\cup S\cup B)&=P(T)+P(S)+P(B)-P(TS)-P(TB)-P(SB)+P(TSB)\\&=\frac{36+28+18-22-12-9+4}{N}=\frac{43}{N}\end{aligned}$$

因此，可以断定有 43 个成员至少参加一种运动. ■

本节的下一个例子不仅提供了一个有点出乎意料的答案，而且还具有理论意义.

例 5m（配对问题） 假定参加某聚会的 N 个人都向房子中央扔出自己的帽子，帽子经充分混合后，每人随机地抽取一顶，试问：

(a) 没有一个人选中自己帽子的概率是多少?

(b) 有 k 个人选中自己帽子的概率是多少?

42

解 (a) 我们首先计算余事件的概率，即至少有一人选中自己帽子的概率. 以 $E_i(i=1,2,\cdots,N)$ 表示第 i 个人选中自己帽子的事件. 由命题 4.4 可得，至少有一人选中自己帽子的概率为

$$P\left(\bigcup_{i=1}^{N}E_i\right)=\sum_{i=1}^{N}P(E_i)-\sum_{i_1<i_2}P(E_{i_1}E_{i_2})+\cdots+(-1)^{n+1}\sum_{i_1<i_2<\cdots<i_n}P(E_{i_1}E_{i_2}\cdots E_{i_n})$$
$$+\cdots+(-1)^{N+1}P(E_1E_2\cdots E_N)$$

如果我们把试验的结果看成由 N 个数组成的一个向量，其中第 i 个分量就是第 i 个人所可能抽取的帽子的个数，则有 $N!$ 种可能的结果.（例如，结果 $(1,2,3,\cdots,N)$ 就表示每个人都选中了自己的帽子.）此外，事件 $E_{i_1}E_{i_2}\cdots E_{i_n}$ 表示有 n 个人 $i_1,i_2,\cdots,i_n$ 选中了自己的帽子，发生的可能方式总共有 $(N-n)(N-n-1)\cdots 3\cdot 2\cdot 1=(N-n)!$ 种. 这是因为，在所剩余的 $N-n$ 个人

中，第一个人可选择 $N-n$ 顶帽子中的任何一顶，第二个人可选择 $N-n-1$ 顶帽子中的任何一顶，等等. 于是，若假定所有 $N!$ 个可能结果是等可能的，则有

$$P(E_{i_1}E_{i_2}\cdots E_{i_n})=\frac{(N-n)!}{N!}$$

再者，由于和式 $\sum_{i_1<i_2<\cdots<i_n}P(E_{i_1}E_{i_2}\cdots E_{i_n})$ 共有 $\binom{N}{n}$ 项，故

$$\sum_{i_1<i_2<\cdots<i_n}P(E_{i_1}E_{i_2}\cdots E_{i_n})=\frac{N!(N-n)!}{(N-n)!n!N!}=\frac{1}{n!}$$

于是有

$$P\left(\bigcup_{i=1}^{N}E_i\right)=1-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{N+1}\frac{1}{N!}$$

因此，没有一个人选中自己帽子的概率为

$$1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+\frac{(-1)^N}{N!}$$

对充分大的 N，这个概率近似等于 $e^{-1}\approx0.367\,88$. 换句话说，当 N 很大时，没有一个人选中自己帽子的概率近似等于 0.37.（多少读者有过“当 $N\to\infty$ 时，此概率趋于 1”这一错误想法?）

(b) 为了得到 N 个人中恰有 k 个人选中自己帽子的概率，我们首先考虑指定的 k 个人选中且只有他们选中自己帽子的种数，这等于其余的 $N-k$ 个人从他们的帽子中没有选中自己帽子的种数，其概率为

43

$$1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+\frac{(-1)^{N-k}}{(N-k)!}$$

因此，所考虑的 k 个人从他们自己的帽子中选中的选法共有

$$(N-k)!\left[1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+\frac{(-1)^{N-k}}{(N-k)!}\right]$$

种. 由于从 N 个人中选取 k 个人又有 $\binom{N}{k}$ 种可能选法，故恰有 k 个人选中自己的帽子共有

$$\binom{N}{k}(N-k)!\left[1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+\frac{(-1)^{N-k}}{(N-k)!}\right]$$

种选法. 因此，所求的概率为

$$\frac{\binom{N}{k}(N-k)!\left[1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+\frac{(-1)^{N-k}}{(N-k)!}\right]}{N!}=\frac{1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+\frac{(-1)^{N-k}}{(N-k)!}}{k!}$$ ⊖

对于很大的 N，它近似等于 $\frac{e^{-1}}{k!}$. 这里得到的概率值 $\frac{e^{-1}}{k!}(k=0,1,2,\cdots)$ 在理论上是很重要的，因为这些值与泊松分布有关系. 这一点将在第 4 章详述. ■

作为命题 4.4 的另一个应用，考虑如下问题.

例 5n 10 对夫妇随机地坐在一张圆桌旁，求没有一个妻子坐在自己丈夫旁边的概率.

⊖ 解这个问题的另一种方法见第 3 章的例 5c.

解　如果以 $E_i(i=1,2,\cdots,10)$ 表示第 i 对夫妇相邻就坐这一事件，那么所求概率为 $1-P\left(\bigcup_{i=1}^{10}E_i\right)$. 现由命题 4.4 得

44

$$P\left(\bigcup_{i=1}^{10}E_i\right)=\sum_{i=1}^{10}P(E_i)-\cdots+(-1)^{n+1}\sum_{i_1<i_2<\cdots<i_n}P(E_{i_1}E_{i_2}\cdots E_{i_n})+\cdots-P(E_1E_2\cdots E_{10})$$

为求 $P(E_{i_1}E_{i_2}\cdots E_{i_n})$，我们首先指出，20 个人围坐在一张圆桌旁共有 19! 种排列方法.（为什么?）使这 n 个指定的男人坐在其妻子旁边的排列个数，可以借助先把这 n 对夫妇中的每一对都看成一个人而方便地得到. 这时，我们需要在圆桌旁安排 $20-n$ 个人，当然应有 $(20-n-1)!$ 种排列. 最后，由于安排每一对夫妇在两个相邻的座位上又有两种可能方法，所以使这 n 个指定的男人与其妻子坐在一起共有 $2^n(20-n-1)!$ 种排列. 因此

$$P(E_{i_1}E_{i_2}\cdots E_{i_n})=\frac{2^n(19-n)!}{(19)!}$$

于是，由命题 4.4，我们得到至少有一对夫妇坐在一起的概率等于

$$\binom{10}{1}2^1\frac{(18)!}{(19)!}-\binom{10}{2}2^2\frac{(17)!}{(19)!}+\binom{10}{3}2^3\frac{(16)!}{(19)!}-\cdots-\binom{10}{10}2^{10}\frac{9!}{(19)!}\approx 0.660\,5$$

而所求的概率近似等于 0.339 5. ■

例 5o（游程问题）　考虑一个运动队以赢 n 次输 m 次的最终结果已经结束赛季. 通过检验赢和输的序列，我们希望确定这个运动队是否有一个比赛的延续期，在这个期间，该运动队比任何期间都最可能赢得比赛. 为了增进对这个问题的理解，在出现 n 次赢和 m 次输的所有 $(n+m)!/(n!\ m!)$ 可能的排列是等可能的假定下，我们计算赢的游程数来看看最有可能出现什么结果. 所谓赢的游程是指赢的序列. 例如，若 $n=10$，$m=6$，比赛的结果序列为 $WWLLWWWLWLLLWWWW$，那么赢的游程数是 4——第一个游程大小是 2，第二个游程大小是 3，第 3 个游程大小是 1，而第 4 个游程大小是 4.

现在设这个运动队赢了 n 次，输了 m 次. 假定所有 $(n+m)!/(n!\ m!)=\binom{n+m}{n}$ 种排列是等可能的，我们来计算赢的游程数正好为 r 的概率. 为此，首先考虑正整数向量 $x_1,x_2,\cdots,x_r$，且 $x_1+x_2+\cdots+x_r=n$，这里 $x_i(i=1,2,\cdots,r)$ 表示第 i 个游程的大小，我们来计算赢的游程数为 r 的结果数. 对任何一个结果，令 y_1 表示第一个赢的游程之前输的次数，y_2 表示第一个与第二个赢的游程之间输的次数，$\cdots$，y_{r+1} 表示最后一个赢的游程之后输的次数，那么 y_i 满足

45

$$y_1+y_2+\cdots+y_{r+1}=m,y_1\geqslant 0,y_{r+1}\geqslant 0,y_i>0,i=2,\cdots,r$$

结果可以形象地表示为

$$\underbrace{LL\cdots L}_{y_1}\ \underbrace{WW\cdots W}_{x_1}\ \underbrace{L\cdots L}_{y_2}\ \underbrace{WW\cdots W}_{x_2}\ \cdots\ \underbrace{WW}_{x_r}\ \underbrace{L\cdots L}_{y_{r+1}}$$

因此，赢的游程数为 r 的结果数（第 i 个游程的大小为 $x_i,i=1,2,\cdots,r$）等于满足上述条件的正整数 $y_1,y_2,\cdots,y_{r+1}$ 的个数，或者等价地，等于正整数

$$\bar{y}_1=y_1+1,\quad \bar{y}_i=y_i,i=2,\cdots,r\quad \bar{y}_{r+1}=y_{r+1}+1$$

的个数，其中

$$\bar{y}_1+\bar{y}_2+\cdots+\bar{y}_{r-1}=m+2$$

由第 1 章命题 6.1 共有 $\binom{m+1}{r}$ 个结果数. 因此赢的游程数为 r 的总结果数是 $\binom{m+1}{r}$ 乘以方程

$x_1+\cdots+x_r=n$ 的正整数解的个数. 再次利用命题 6.1 得，赢的游程数为 r 的结果总数是 $\binom{m+1}{r}\binom{n-1}{r-1}$. 因为有 $\binom{m+n}{n}$ 个等可能的结果，故所求概率为

$$P(\{赢的游程数为\ r\})=\frac{\binom{m+1}{r}\binom{n-1}{r-1}}{\binom{m+n}{n}} \qquad r\geqslant 1$$

例如，若 $n=8$，$m=6$，所有 $\binom{14}{8}$ 个结果是等可能的，则赢的游程数为 7 的概率是 $\dfrac{\binom{7}{7}\binom{7}{6}}{\binom{14}{8}}$ $=1/429$. 因此，若比赛的结果为 $WLWLWLWLWWLWLW$，那么我们可能猜测该队赢的概率整个会发生变化.（特别地，当该队输了它的上一次比赛时，该队赢的概率似乎比较高，而当它赢了上一场比赛时，赢的概率相当低.）考虑一种极端情况，如果比赛的结果是 $WWWWWWWWLLLLLL$，那么赢的游程数为 1，而概率为 $P(\{赢的游程数为\ 1\})=\dfrac{\binom{7}{1}\binom{7}{0}}{\binom{14}{8}}=\dfrac{1}{429}$，因此，

46

该队赢得比赛的概率在整个 14 场比赛中保持不变. ■

*2.6 概率作为一种连续的集函数

如果

$$E_1\subset E_2\subset\cdots\subset E_n\subset E_{n+1}\subset\cdots$$

则称事件序列 $\{E_n,n\geqslant 1\}$ 为增序列. 如果

$$E_1\supset E_2\supset\cdots\supset E_n\supset E_{n+1}\supset\cdots$$

则称事件序列 $\{E_n,n\geqslant 1\}$ 为减序列.

设 $\{E_n,n\geqslant 1\}$ 是一个增的事件序列，我们定义一个新的事件，记为 $\lim\limits_{n\to\infty}E_n$，由

$$\lim_{n\to\infty}E_n=\bigcup_{i=1}^{\infty}E_i$$

给出. 类似地，设 $\{E_n,n\geqslant 1\}$ 是一个减的事件序列，我们用

$$\lim_{n\to\infty}E_n=\bigcap_{i=1}^{\infty}E_i$$

定义事件 $\lim\limits_{n\to\infty}E_n$.

我们现在证明命题 6.1.

命题 6.1 如果 $\{E_n,n\geqslant 1\}$ 是一个增的事件序列或是一个减的事件序列，则

$$\lim_{n\to\infty}P(E_n)=P(\lim_{n\to\infty}E_n)$$

证明 首先假定 $\{E_n,n\geqslant 1\}$ 是一个增的事件序列，定义事件 $F_n(n\geqslant 1)$ 为

$$F_1=E_1$$

$$F_n=E_n\left(\bigcup_{1}^{n-1}E_i\right)^c=E_nE_{n-1}^c \qquad n>1$$

由于事件序列是增的，所以上面我们利用了结果 $\bigcup_{1}^{n-1} E_i = E_{n-1}$. 换句话说，$F_n$ 由在 E_n 中而不在任何 $E_i(i < n)$ 中的结果所组成. 容易验证 F_n 是互不相容的事件，并且

$$\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i \qquad \text{和} \qquad \bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i \qquad \text{对一切 } n \geqslant 1$$

47

因此

$$\begin{aligned} P\Big(\bigcup_{i=1}^{\infty} E_i\Big) &= P\Big(\bigcup_{i=1}^{\infty} F_i\Big) \\ &= \sum_{i=1}^{\infty} P(F_i) \quad (\text{由公理 } 3) \\ &= \lim_{n\to\infty}\sum_{i=1}^{n} P(F_i) = \lim_{n\to\infty} P\Big(\bigcup_{i=1}^{n} F_i\Big) \\ &= \lim_{n\to\infty} P\Big(\bigcup_{i=1}^{n} E_i\Big) = \lim_{n\to\infty} P(E_n) \end{aligned}$$

当 $\{E_n, n\geqslant 1\}$ 是增序列时，结论得证.

若 $\{E_n, n\geqslant 1\}$ 是一个减的事件序列，则 $\{E_n^c, n\geqslant 1\}$ 是一个增序列，因此由上述方程有

$$P\Big(\bigcup_{1}^{\infty} E_i^c\Big) = \lim_{n\to\infty} P(E_n^c)$$

但是由于 $\bigcup_{1}^{\infty} E_i^c = \Big(\bigcap_{1}^{\infty} E_i\Big)^c$，所以我们有

$$P\Big(\Big(\bigcap_{1}^{\infty} E_i\Big)^c\Big) = \lim_{n\to\infty} P(E_n^c)$$

或者等价地，

$$1 - P\Big(\bigcap_{1}^{\infty} E_i\Big) = \lim_{n\to\infty}[1 - P(E_n)] = 1 - \lim_{n\to\infty} P(E_n)$$

或

$$P\Big(\bigcap_{1}^{\infty} E_i\Big) = \lim_{n\to\infty} P(E_n)$$

结论得证. ■

例 6a（概率和一个悖论） 假定我们有一个无限大的罐和标号分别为 1，2，3，… 的球的集合. 考虑如下试验，在下午差 1 分 12 点到 12 点将 1～10 号球放入罐中，并将 10 号球抽出（假定抽球不花时间）. 在下午差 $\frac{1}{2}$ 分 12 点到 12 点将 11～20 号球放入罐中，并将 20 号球抽出. 在下午差 $\frac{1}{4}$ 分 12 点到 12 点将 21～30 号球放入罐中，并将 30 号球抽出. 在下午差 $\frac{1}{8}$ 分 12 点到 12 点将 31～40 号球放入罐中，并将 40 号球抽出，如此继续下去，我们感兴趣的问题是，12 点时罐中共有多少个球？

48

问题的答案是很清楚的，在 12 点时，罐中有无穷多个球，因为标号不是 $10n(n\geqslant 1)$ 的任意一个球在 12 点前被放入罐中并且没有被抽出. 因此，当试验按上述方式进行时问题得到解决.

然而，我们现在改变试验的方式，在下午差 1 分 12 点到 12 点将 1～10 号球放入罐中，并将 1

号球抽出；在下午差$\frac{1}{2}$分 12 点到 12 点将 11～20 号球放入罐中，并将 2 号球抽出；在下午差$\frac{1}{4}$分 12 点到 12 点将 21～30 号球放入罐中，并将 3 号球抽出；在下午差$\frac{1}{8}$分 12 点到 12 点将 31～40 号球放入罐中，并将 4 号球抽出；如此继续下去. 对这个新的试验，12 点时罐中共有多少个球？

我们十分惊奇地发现，答案是 12 点时罐子是空的. 因为考虑任意一个球，比如说标号为 n 的球，在到 12 点的某一时间段（特别是，在下午差$\left(\frac{1}{2}\right)^{n-1}$分 12 点到 12 点之间），这个球已经被从罐中取出. 因此，对每个 n，标号为 n 的球在 12 点时不在罐中，所以 12 点时罐必为空.

通过前面的讨论我们看到，不同的抽球方式得到不同的结论. 在第一种情况下，只有标号为 $10n(n\geqslant 1)$ 的球被抽出，而在第二种情况下，最终所有的球被抽出. 现在我们假定从这些球中随机地选一个球，并将其抽出. 即假定在差 1 分 12 点到 12 点之间将标号为 1～10 的球放入罐中，并从中随机地选一个球将其抽出来，如此继续下去，在这种情况下，问 12 点时罐中有多少个球？

解 我们现在证明在 12 点时以概率 1 罐是空的. 首先考虑标号为 1 的球. 令 E_n 表示经过 n 次抽球后 1 号球还在罐中的事件. 显然有

$$P(E_n)=\frac{9\cdot 18\cdot 27\cdots(9n)}{10\cdot 19\cdot 28\cdots(9n+1)}$$

（为了理解这个等式，注意假如第 n 次抽球后 1 号球仍在罐中，那么第一个被抽出的球可以是另外 9 个球中的任何一个，第二个被抽出的球可以是另外 18 个球中的任何一个（在第二次抽球时，罐中有 19 个球，其中的一个必须是 1 号球），如此等等. 以同样的方式可得到分母的值. ）

49

现在，标号为 1 的球在 12 点时仍在罐中的事件恰好是事件 $\bigcap_{n=1}^{\infty} E_n$. 因为事件 $E_n(n\geqslant 1)$ 是减序列，因此由命题 6.1 得到

$$P\{12\text{ 点时 1 号球在罐中}\}=P\left(\bigcap_{n=1}^{\infty} E_n\right)=\lim_{n\to\infty}P(E_n)=\prod_{n=1}^{\infty}\left(\frac{9n}{9n+1}\right)$$

我们现在证明

$$\prod_{n=1}^{\infty}\frac{9n}{9n+1}=0$$

因为

$$\prod_{n=1}^{\infty}\left(\frac{9n}{9n+1}\right)=\left[\prod_{n=1}^{\infty}\left(\frac{9n+1}{9n}\right)\right]^{-1}$$

所以这等价于证明

$$\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right)=\infty$$

对一切 $m\geqslant 1$，

$$\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right)\geqslant\prod_{n=1}^{m}\left(1+\frac{1}{9n}\right)=\left(1+\frac{1}{9}\right)\left(1+\frac{1}{18}\right)\left(1+\frac{1}{27}\right)\cdots\left(1+\frac{1}{9m}\right)$$

$$>\frac{1}{9}+\frac{1}{18}+\frac{1}{27}+\cdots+\frac{1}{9m}=\frac{1}{9}\sum_{i=1}^{m}\frac{1}{i}$$

因此，令 $m\to\infty$ 并利用 $\sum_{i=1}^{\infty}\frac{1}{i}=\infty$ 得到

$$\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right)=\infty$$

令 F_i 表示12点时标号为 i 的球在罐中的事件，我们已证明了 $P(F_1)=0$. 同理可以证明对一切 i，$P(F_i)=0$.（例如，同理可证 $P(F_i)=\prod_{n=2}^{\infty}[9n/(9n+1)]$，$i=11,12,\cdots,20$.）因此，在12点时，罐不空的概率为 $P\left(\bigcup_{1}^{\infty}F_i\right)$，由布尔不等式(见自测题与练习14)，有 50

$$P\left(\bigcup_{1}^{\infty}F_i\right)\leqslant\sum_{1}^{\infty}P(F_i)=0$$

于是，在12点时以概率1罐是空的. ■

2.7 概率作为一种置信的度量

至此，我们解释了在试验连续重复进行的情形下，给定试验的事件的概率作为事件出现的频率的一个度量. 然而，还有另外一些概率术语的使用场合. 例如，我们听到过这样一些说法，“莎士比亚以90%的可能性写了《哈姆雷特》”或者“奥斯瓦德暗杀肯尼迪的概率只有80%”. 我们如何解释这些说法呢?

最简单而又最自然的解释是概率作为人们对观点相信程度的一个度量. 换句话说，上述观点的人比较肯定奥斯瓦德暗杀肯尼迪且十分肯定莎士比亚写了《哈姆雷特》. 作为人们相信程度的度量的概率的解释常常被认为是个人的或者主观的概率.

假定“置信的度量”满足概率论的所有公理是符合逻辑的. 例如，我们以70%的概率认为莎士比亚写了《凯撒大帝》、以10%的概率认为是马洛写的，那么从逻辑上来讲我们以80%的概率认为莎士比亚或者马洛写了《凯撒大帝》. 因此，我们解释概率作为一种置信的度量或者作为一种长久出现的频率，其数学性质保持不变.

例7a 假定在7匹马的比赛中，你认为前两匹马赢得比赛的机会各为20%，第3、4匹马赢的机会各为15%，余下的3匹马赢的机会各为10%. 以多少钱下赌注前三匹马中的一匹赢得比赛，又以多少钱下赌注1、5、6、7匹马中的一匹为获胜者? 哪一种赌注对你来说是比较好的?

解 根据你关于比赛的置信概率，第一种赌注赢的概率为0.2+0.2+0.15=0.55，第二种赌注赢的概率为0.2+0.1+0.1+0.1=0.5. 因此，第一种赌注是比较诱人的. ■ 51

在假设中，应该注意到人们的主观概率总是与概率论的公理是一致的，我们论述的是一个理想的人而不是一个实际的人. 例如，若我们问某人他认为下列事件的机会是多少:

(a) 今天下雨.

(b) 明天下雨.

(c) 今天和明天都下雨.

(d) 今天或明天下雨.

经过一段深思熟虑，这个人可能给出30%、40%、20%和60%作为回答. 然而这样的回答(或者这样的主观概率)与概率论的公理是不一致的.（为什么?）当然我们通过指出这种回答存在的问题希望他改变其答案.（我们可接受的概率分别为30%、40%、10%和60%.）

小结

用 S 表示一个试验所有可能结果的集合，称 S 为该试验的样本空间. 一个事件就是 S 的一

个子集．若 $A_i(i=1,2,\cdots,n)$ 是事件，则 $\bigcup_{i=1}^{n} A_i$ 称为这些事件的并，它由在事件 $A_i(i=1,\cdots,n)$ 中的至少一个事件的所有结果组成．同样，$\bigcap_{i=1}^{n} A_i$（有时写为 $A_1A_2\cdots A_n$）称为事件 A_i 的交，它由在所有事件 $A_i(i=1,\cdots,n)$ 中的所有结果组成．

对任意事件 A，我们定义 A^c 为在样本空间而不在 A 中的所有结果组成的事件，A^c 称为事件 A 的余事件．事件 S^c 的试验结果是空的，记为 $\varnothing$，并称为空集．如果 $AB=\varnothing$，那么我们说事件 A 与 B 互不相容．

对样本空间 S 的每一个事件 A，我们称满足以下条件的数 $P(A)$ 为事件 A 的概率：

(i) $0\leqslant P(A)\leqslant 1$.

(ii) $P(S)=1$.

(iii) 对互不相容的事件 $A_i(i\geqslant 1)$，

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i)$$

$P(A)$ 表示在 A 中的试验结果的概率．

容易证明

$$P(A^c)=1-P(A)$$

一个有用的结果是

52

$$P(A\cup B)=P(A)+P(B)-P(AB)$$

这个公式可一般化为

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)-\sum_{i<j}\sum P(A_iA_j)+\sum_{i<j<k}\sum\sum P(A_iA_jA_k)+\cdots+(-1)^{n+1}P(A_1A_2\cdots A_n)$$

如果 S 是有限集，每个点集假定具有相等的概率，则

$$P(A)=\frac{|A|}{|S|}$$

这里 $|E|$ 表示在事件 E 中的结果数．

$P(A)$ 既可以解释为一种长久出现的相对频率又可以解释为一种置信的度量．

习题

1. 盒中有三粒弹子，一红一绿一蓝．考虑这样一个试验：从盒中拿出一粒弹子，然后把它放回并取出第二粒弹子，试写出样本空间．若取出的第一粒弹子不放回就取出第二粒弹子，样本空间该怎样描述？
2. 连续掷一个骰子直到6点出现时试验停止．此试验的样本空间是什么？令 E_n 表示完成此试验需要掷 n 次这一事件，样本空间中哪些点属于 E_n？$\left(\bigcup_{n=1}^{\infty} E_n\right)^c$ 是怎样一个事件？
3. 掷2个骰子，以 E 表示它们的点数之和为奇数这一事件，F 表示至少一个骰子是1点的事件，而 G 表示点数之和是5这一事件．事件 $EF, E\cup F, FG, EF^c, EFG$ 各表示什么？
4. A，B 和 C 轮流抛一枚硬币，A 先抛，然后 B 抛，然后 C 抛，然后 A 抛，规定首先抛出正面者获胜．此试验的样本空间可写作

$$S=\begin{cases}1,01,001,0001,\cdots,\\0000\cdots\end{cases}$$

(a) 试解释此样本空间.

(b) 根据 S 定义如下事件：

(i) $A=A$ 获胜. (ii) $B=B$ 获胜. (iii) $(A\cup B)^c$.

5. 某系统由 5 个元件组成，每一个元件要么工作要么失效. 考虑一个观察每一个元件的状态组成的试验，其试验的结果用向量(x_1,x_2,x_3,x_4,x_5)表示，这里 x_i 等于 1 表示第 i 个元件工作，x_i 等于 0 表示第 i 个元件失效.

(a) 这个试验的样本空间中有多少个结果?

(b) 假定当元件 1 和元件 2 都正常工作，或元件 3 和元件 4 都正常工作，或元件 1、3 和 5 都正常工作时，该系统正常工作. 令 W 表示系统能正常工作的事件，写出 W 的所有结果.

53

(c) 令 A 表示元件 4 和元件 5 都失效的事件. 事件 A 包含了多少个结果?

(d) 写出事件 AW 的所有结果.

6. 某医院的管理者对遭受枪击伤害的入院病人根据他们是否投了保险进行编码(编码 1 对应投保，0 对应未投保)，根据他们的伤势情况又分为良好(g)、尚可(f)、严重(s). 考虑这样一个病人的编码组成的试验.

(a) 写出该试验的样本空间.

(b) 令 A 表示病人处于严重状态的事件，写出事件 A 的结果.

(c) 令 B 表示病人未投保的事件，写出 B 的结果.

(d) 给出事件 $B^c\cup A$ 的所有结果.

7. 考虑一个成人足球队的 15 个队员的工作类型(蓝领或者白领)和政治团体(共和党、民主党或者无党派)所组成的试验. 下列事件有多少结果?

(a) 抽样空间.

(b) 足球队中至少有一个队员是蓝领工人的事件.

(c) 球队中没有一个队员考虑自己是无党派人士的事件.

8. 假定事件 A 和 B 是互不相容的事件，且 $P(A)=0.3$, $P(B)=0.5$，下列事件的概率是多少?

(a) 事件 A 或 B 出现.

(b) 事件 A 出现而事件 B 不出现.

(c) 事件 A 和事件 B 同时出现.

9. 某零售公司接受美国万国宝通银行卡或者 VISA 信用卡. 该公司的顾客中 24%携带美国万国宝通银行卡，61%携带 VISA 卡，11%携带这两种卡. 该公司的顾客中只携带一种卡使得零售公司能接受的百分比是多少?

10. 某学校 60%的学生既不戴戒指又不戴项链，20%的学生戴戒指，30%的学生戴项链. 随机地从该校中选择一名学生，计算下列事件的概率：

(a) 该学生戴戒指或戴项链. (b) 该学生既戴戒指又戴项链.

11. 美国男人中的 28%抽香烟，7%抽雪茄，而 5%既抽香烟又抽雪茄.

(a) 美国男人中既不抽香烟又不抽雪茄的百分比是多少?

(b) 美国男人中抽雪茄而不抽香烟的百分比是多少?

12. 某小学开设 3 种语言班：西班牙语、法语和德语. 这些语言班对该校的任意 100 名学生开放. 有 28 名学生在西班牙语班，26 名学生在法语班，16 名学生在德语班. 有 12 名学生既在西班牙语班又在法语班，4 名学生既在西班牙语班又在德语班，6 名学生既在法语班又在德语班，有 2 名学生 3 个班都参加.

(a) 若随机地选择一名学生，他不在任何一个班的概率是多少?

(b) 若随机地选择一名学生，他恰好参加一个语言班的概率是多少?

(c) 若随机地选择2名学生，至少有一名学生参加一个语言班的概率是多少？

54

13. 某10万人口的城市中有Ⅰ、Ⅱ与Ⅲ三份报纸，此城市居民中读上述报纸的比例如下：

Ⅰ：10%　　Ⅰ与Ⅱ：8%　　Ⅰ、Ⅱ与Ⅲ：1%

Ⅱ：30%　　Ⅰ与Ⅲ：2%

Ⅲ：5%　　Ⅱ与Ⅲ：4%

(此表告诉我们，读Ⅰ与Ⅱ两种报纸的有8 000人.)

(a) 求只读一种报纸的人数.

(b) 有多少人读至少两种报纸？

(c) 若Ⅰ与Ⅲ是晨报，Ⅱ是晚报，有多少人读至少一种晨报加上一种晚报？

(d) 有多少人不读任何报纸？

(e) 有多少人只读一种晨报加上一种晚报？

14. 对某杂志的1 000名订阅者进行一次调查，得到如下数据：就工作、婚姻与教育状况而言，有312个专业人员，470个已婚人员，525个大学毕业生，42个专业院校毕业生，147个已婚大学毕业生，86个已婚专业人员，25个已婚专业院校毕业生. 试说明这些调查数据一定是错误的.

提示：令M,W和G分别表示专业人员、已婚人员和大学毕业生的集合，假定随机地从这1 000人中随机地选择一人，可用命题4.4证明，如果上述数据是正确的，那么有$P(M\cup W\cup G)>1$.

15. 如果假定所有$\binom{52}{5}$种发纸牌的结果是等可能的，试求发到如下牌的概率.

(a) 同花大顺(如果所有五张牌是同一花色，则称这手牌为同花大顺).

(b) 一对(正好两张点数相同而另三张点数不同，即a,a,b,c,d，其中a,b,c,d各不相同).

(c) 两对(正好有两个两张同点而另一张是另外的点，即a,a,b,b,c，其中a,b,c各不相同).

(d) 正好有三张点数相同而另两张点数不同(即a,a,a,b,c，其中a,b,c各不相同).

(e) 正好四张点数相同而另一张点数不同(即a,a,a,a,b，其中a,b不相同).

16. 同时掷5个骰子，试证：

(a) $P\{$点数各不相同$\}=0.092\,6$.　　(b) $P\{$一对$\}=0.463\,0$.

(c) $P\{$两对$\}=0.231\,5$.　　(d) $P\{$恰有三个点数相同$\}=0.154\,3$.

(e) $P\{$满堂$\}=0.038\,6$.　　(f) $P\{$恰有4个点数相同$\}=0.019\,3$.

(g) $P\{$点数全同$\}=0.000\,8$.

17. 将8个“车”随机地放在国际象棋的棋盘上，计算没有任何一个车能吃掉其他车的概率，即求棋盘上各行各列都至多放着一个车的概率.

18. 从一副纸牌中随机地抽出两张. 问正好抽到21点的概率是多少？即，正好抽到一张是A，而另一张是10，J，Q，K中的某一张的概率是多少？

19. 两个匀称的骰子有两面染成红色，两面染成黑色，一面染成黄色，一面染成白色. 同时掷这两个骰子，同一颜色的面出现的概率是多少？

20. 假定你与庄家玩21点，在一次洗牌后，你和庄家都没有发到21点的概率是多少？

21. 一个小的社区组织由20个家庭组成，其中4个家庭有1个孩子，8个家庭有2个孩子，5个家庭有3个孩子，2个家庭有4个孩子，1个家庭有5个孩子.

55

(a) 假如从这些家庭中随机地选择一个家庭，其有$i(i=1,2,3,4,5)$个孩子的概率是多少？

(b) 假如随机地选择一个孩子，这个孩子是来自有$i(i=1,2,3,4,5)$个孩子的家庭的概率是多少？

22. 用如下方法洗有n张纸牌的一副牌：对这副牌的任意的最初次序，对每张牌每次通过抛一枚硬币换一张牌. 假如硬币出现正面，这张牌仍留在原来位置，假如硬币出现反面，那么将这张牌放到这副牌的最后位置. 在抛了n次硬币以后，我们就说完成了一轮. 例如，假如$n=4$，最初的次序是1，2，3，4，

若硬币抛的结果是 h,t,t,h，那么一轮结束时，该副牌的次序是 1，4，2，3. 假定 n 个硬币抛的序列的所有结果是等可能的，一轮以后纸牌的次序与原来的次序相同的概率是多少？

23. 掷两个匀称的骰子，第二个骰子的点数大于第一个骰子的点数的概率是多少？
24. 掷两个骰子，问掷出的点数之和为 i 的概率是多少？试对 $i=2,3,\cdots,11,12$ 计算此概率.
25. 连续掷一对骰子，直到点数之和是 5 或 7 为止. 试求 5 点出现的概率.
    提示：令 E_n 表示第 n 次掷出 5 点，而在前 $n-1$ 次 5 点、7 点都没有出现这一事件. 求出 $P(E_n)$，并说明 $\sum_{n=1}^{\infty} P(E_n)$ 就是所求的概率.
26. 双骰子游戏这样进行：玩家掷两个骰子，如果点数之和为 2，3 或 12，他就输了；如果点数之和为 7 或 11，他就赢了；如果掷出其他点，就接着再掷，直到掷出第一次出现的点数或 7 点为止. 若先掷出 7，他输掉；若在 7 点之前他再次掷出第一次的点数，则他赢. 试求这个玩家赢的概率.
    提示：令 E_i 表示第一次掷出 i 点且玩家赢这一事件. 所求的概率是 $\sum_{i=2}^{12} P(E_i)$. 为计算 $P(E_i)$，再次令 $E_{i,n}$ 表示第一次掷出 i 点而玩家在掷第 n 次时赢这一事件，证明 $P(E_i)=\sum_{n=1}^{\infty} P(E_{i,n})$.
27. 罐中有 3 个红球和 7 个黑球，A、B 两人轮流从罐中取球，A 先取一个，然后 B，然后 A，取出后不放回，如此继续直到取出一个红球为止，求 A 先取到红球的概率.
28. 罐中有 5 个红球、6 个蓝球和 8 个绿球. 现随机地取出 3 个球，计算下列事件的概率：
    (a) 所取的球颜色相同. (b) 所取的球颜色不同.
    如果假设每个球取出后总是记住其颜色，并在取下一个球之前将它放回罐中，上述两个事件的概率是多少？这称为*有放回抽样*.
29. 一个箱中有 n 个白球和 m 个黑球，这里 n 和 m 均为正整数.
    (a) 假如不放回地随机地抽取 2 个球，2 个球颜色相同的概率是多少？
    (b) 假如有放回地随机地抽取 2 个球，2 个球颜色相同的概率是多少？
    (c) 证明(b)的概率总是大于(a)的概率.

56

30. 两个学校的国际象棋俱乐部分别有 8 名和 9 名棋手，从每个俱乐部随机地选择 4 名棋手参加两个学校之间的比赛. 从一个队选择的棋手随机地与另外一个队的棋手进行配对，每对下一盘棋. 假定 Rebecca 和她的妹妹 Elise 属于不同学校的象棋俱乐部. 计算下列事件的概率：
    (a) Rebecca 和 Elise 将被配对.
    (b) Rebecca 和 Elise 分别代表各自的学校参加比赛但不相互配对.
    (c) Rebecca 和 Elise 恰有一人被选上代表她的学校参赛.
31. 三人篮球队由一后卫、一中锋和一前锋组成.
    (a) 现从 3 个不同的三人篮球队中各随机地选一人，问正好组成一个完整球队的概率是多少？
    (b) 被选出的 3 人都打同一个位置的概率是多少？
32. 由 b 个男孩和 g 个女孩组成的群体随机地排成一列，假定 $(b+g)!$ 个排列中的每一个排列是等可能的，问第 $i(1\leqslant i\leqslant b+g)$ 个位置是女孩的概率是多少？
33. 一个森林中有 20 只麋鹿，捕获了其中的 5 只，做了标记后将其放归森林. 过了一段时间，捕获了其中的 4 只. 问这 4 只中有 2 只带有记号的概率是多少？你进行了什么假定？
34. 据报道，亚伯洛的二等伯爵有一个输 1 000 赢 1 的赌注：13 张牌的一手桥牌中至少包含一张大于等于 10 的牌.（大于等于 10 是指这张牌是 10，J，Q，K 或者 A.）现在，我们称没有大于 9 的一手牌为 Yarborough. 问随机地选择一手牌正好为 Yarborough 的概率是多少？
35. 有 30 位精神病医生和 24 位心理学家参加某一会议. 从这 54 人中随机地选择 3 人参加专门小组讨论. 问至少选择到一位心理学家的概率是多少？

36. 从 52 张牌中随机地挑选 2 张，计算下列事件的概率：

(a) 两张均为 A. (b) 两张的数字相同.

37. 某位导师给她班上的学生出了 10 道题，并且告诉他们从中随机地选择 5 道题作为最终的考试题. 假如一个学生已经会做其中的 7 道题，试计算下列事件的概率：

(a) 全部 5 题回答正确. (b) 至少有 4 题回答正确.

38. 抽屉中有 n 双短袜，其中 3 双是红色的. 从中随机地选择 2 双，当 n 等于多少时，所选择的 2 双均为红色的概率是 $\frac{1}{2}$?

39. 某城市有 5 家旅馆，有一天，有 3 人去该城市里的旅馆投宿，问 3 人分住在三处的概率是多少？对此题你进行了怎样的假定？

40. 某城市有 4 名电视修理工，如果有 4 台电视机坏了，问正好有 i 名修理工被找的概率是多少？对 $i=1,2,3,4$ 解此题. 你要进行怎样的假设？

41. 一个骰子被掷了 4 次，6 点至少出现一次的概率是多少？

42. 两个骰子接连被掷了 n 次，求至少出现一个双 6 点的概率. 问需多大的 n，才能使此概率不小于 $\frac{1}{2}$?

43. (a) 包括 A、B 在内的 N 个人随机地排成一行，A 与 B 彼此相邻的概率是多少？

57

(b) 假如这 N 个人随机地站成一圈，上述概率是多少？

44. A,B,C,D,E 5 个人排成一列，假定每一种可能的排列是等可能的，求下列事件的概率：

(a) 在 A 和 B 之间恰好有一个人.

(b) 在 A,B 之间恰好有两个人.

(c) 在 A,B 之间有 3 个人.

45. 一位妇女有 n 把钥匙，其中只有一把能打开她的门.

(a) 假如她随机地选一把钥匙试开，打不开的就丢掉，问她第 k 次试开时打开门的概率是多少？

(b) 如果试过的钥匙不丢掉，上述概率又是多少？

46. 为了使同一个房间的人中至少有两个人在同一个月过生日的概率至少是 $\frac{1}{2}$，问该房间中应该有多少人？假定一年的 12 个月是等可能的.

47. 若在一个房间中有 12 个陌生人，问没有两个人在同一个月过生日的概率是多少？

48. 给定 20 个人，问在 12 个月中正好 2 个人过生日的月份有 4 个、3 个人过生日的月份有 4 个的概率是多少？

49. 一群人由 6 个男人 6 个女人组成，现随机地把他们分成两组，问每组各有 3 个男人的概率是多少？

50. 打桥牌时，试求你手中有 5 张黑桃，而另外 8 张黑桃全在你的同伴手中的概率.

51. 将 n 个球随机地放在 N 个格子中，假定所有 N^n 种放法是等可能的，求有 m 个球放在第一个格子中的概率.

52. 柜子里有 10 双鞋. 现随机地取出 8 只，试求下列事件的概率：

(a) 取出的鞋都不成双. (b) 取出的鞋恰好成一双.

53. 4 对夫妇排成一行，求没有一个丈夫与他的妻子相邻的概率.

54. 试计算一手桥牌中至少缺一种花色的概率. 注意，答案不是

$$\frac{\binom{4}{1}\binom{39}{13}}{\binom{52}{13}}$$

为什么？

提示：用命题 4.4.

55. 在 13 张牌的一手牌中计算下列事件的概率：
   (a) 至少包含一种花色的 A 和 K.
   (b) 包含 13 个点中的至少一个点的全部 4 张牌.

56. 两个人玩如下游戏：A 在图 2-6 的转盘中选择一个，然后 B 在剩下的两个转盘中选一个，两个人同时转动各自的转盘，转盘停止时，数字大的一方为胜者. 假定每个转盘 3 个区域中的任何一个区域停止是等可能的，你认为 A 获胜还是 B 获胜？试解释你的回答！ 58

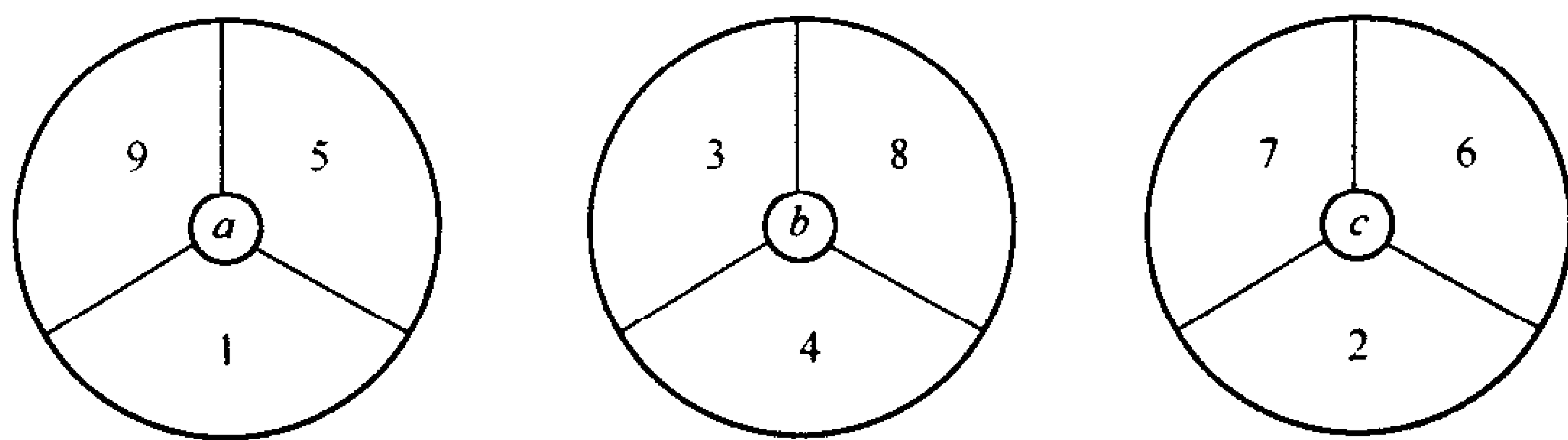


图 2-6

理论练习

证明下列结论(1～4).

1. $EF \subset E \subset E \cup F$.
2. 若 $E \subset F$，则 $F^c \subset E^c$.
3. $F = FE \cup FE^c$，$E \cup F = E \cup E^cF$.
4. $\left(\bigcup_1^\infty E_i\right)F = \bigcup_1^\infty E_iF$，$\left(\bigcap_1^\infty E_i\right) \cup F = \bigcap_1^\infty (E_i \cup F)$.
5. 对任何事件序列 $E_1, E_2, \cdots$，试定义一个新的两两不交的事件序列 $F_1, F_2, \cdots$(即当 $i \neq j$ 时，总有 $F_iF_j = \varnothing$)，使得对一切 $n \geqslant 1$，有
$$\bigcup_1^n F_i = \bigcup_1^n E_i$$
6. 设 E, F, G 为三个事件，试将下列各事件用 E, F 与 G 表示出来.
   (a) 只有 E 发生. (b) E 与 G 都发生但 F 不发生.
   (c) 至少一个事件发生. (d) 至少两个事件发生.
   (e) 三个事件都发生. (f) 三个事件都不发生.
   (g) 至多一个事件发生. (h) 至多两个事件发生.
   (i) 正好两个事件发生. (j) 至多三个事件发生.
7. 化简下列各式.
   (a) $(E \cup F)(E \cup F^c)$. (b) $(E \cup F)(E^c \cup F)(E \cup F^c)$.
   (c) $(E \cup F)(F \cup G)$. 59
8. 令 S 是一个给定的集合，假如对某个 $k > 0$，$S_1, S_2, \cdots, S_k$ 是 S 的互不相交的非空子集，满足 $\bigcup_{i=1}^k S_i = S$，那么就称集合 $\{S_1, S_2, \cdots, S_k\}$ 为 S 的一个分割. 令 T_n 表示集合 $\{1, 2, \cdots, n\}$ 的不同分割的个数. 因此，$T_1 = 1$(只有一个分割是 $S_1 = \{1\}$)，$T_2 = 2$(两个分割是 $\{\{1, 2\}\}$，$\{\{1\}, \{2\}\}$).
   (a) 通过计算所有分割，证明 $T_3 = 5, T_4 = 15$.
   (b) 证明

$$T_{n+1} = 1 + \sum_{k=1}^{n} \binom{n}{k} T_k$$

并利用上式计算 T_{10}.

提示：选择 $n+1$ 项的一个分割的方法是将其中的一项固定. 首先选择 $k(k=0,1,\cdots,n)$，我们得到不同的分割，然后对非固定的大小为 $n-k$ 的子集，得到不同的分割，最后对剩余的 k 个非固定的项，得到任意的 T_k 个分割. 通过添加固定的项到大小为 $n-k$ 的子集上我们得到所有 $n+1$ 项的分割数.

9. 假设某试验进行 n 次，对于其样本空间中的任一事件 E，令 $n(E)$ 表示事件 E 发生的次数，并定义 $f(E)=\frac{n(E)}{n}$，试证 $f(\cdot)$满足公理 1，2，3.

10. 证明：
$$P(E\cup F\cup G)=P(E)+P(F)+P(G)-P(E^cFG)-P(EF^cG)-P(EFG^c)-2P(EFG)$$

11. 若 $P(E)=0.9$，而 $P(F)=0.8$，试证 $P(EF)\geqslant 0.7$. 一般地，也就是证明如下的 Bonferroni 不等式
$$P(EF)\geqslant P(E)+P(F)-1$$

12. 证明事件 E 与 F 恰好有一个发生的概率为 $P(E)+P(F)-2P(EF)$.

13. 试证 $P(EF^c)=P(E)-P(EF)$.

14. 用数学归纳法证明命题 4.4.

15. 罐中有 M 个白球和 N 个黑球，现随机地取出 r 个球，问正好包含 k 个白球的概率是多少？

16. 用归纳法将 Bonferroni 不等式推广到 n 个事件的情形，即证明
$$P(E_1E_2\cdots E_n)\geqslant P(E_1)+P(E_2)+\cdots+P(E_n)-(n-1)$$

17. 考虑例 5m 所述的配对问题. 令 A_N 表示“N 个人在他们的 N 顶帽子中任选一顶都没有选中自己帽子”的种数，试论证
$$A_N=(N-1)(A_{N-1}+A_{N-2})$$
此时加上边界条件 $A_1=0,A_2=1$ 之后可解出 A_N，并且所求的没有一个人选中自己帽子的概率为$\frac{A_N}{N!}$.

提示：在第一个人选取了某一顶别人的帽子之后，其余 $N-1$ 个人是在 $N-1$ 顶帽子中选取，但所剩的帽子中必然不包含这 $N-1$ 个人中某人的帽子. 因此，这时有一个额外的人和一顶额外的帽子. 于是，没有一个人选中自己帽子的问题可以分为额外的人选中额外的帽子与额外的人没有选中额外的帽子两种情况来讨论.

18. 将一枚硬币抛 n 次，令 f_n 表示正面总不连续出现的抛法种数，试论证
$$f_n=f_{n-1}+f_{n-2} \qquad n\geqslant 2，其中\ f_0\equiv 1, f_1\equiv 2$$

60

提示：正面开始有多少个结果，反面开始有多少个结果？

如果以 P_n 表示抛一枚硬币 n 次正面不连续出现的概率，假定 n 次试验的所有可能结果是等可能的，试用 f_n 表示 P_n. 算出 P_{10}.

19. 某箱中有 n 个红球和 m 个蓝球，每次抽一球，直到抽出 $r(r\leqslant n)$个红球为止，求总共抽出 k 个球的概率.

提示：总共抽出 k 个球当且仅当前 $k-1$ 次抽了 $r-1$ 个红球，而第 k 次抽球时抽到红球.

20. 若一个试验的样本空间由可数无穷多个点组成，试说明所有的点不能是等可能的. 各点发生的概率能不能都是正的？

*21. 当 n 次赢和 m 次输是随机变化时，例 5o 考虑了赢的游程数的概率. 现在我们考虑游程总数——赢的游程数加上输的游程数，证明
$$P\{游程数为\ 2k\}=2\frac{\binom{m-1}{k-1}\binom{n-1}{k-1}}{\binom{m+n}{n}}$$

$$P\{\text{游程数为 } 2k+1\}=\frac{\binom{m-1}{k-1}\binom{n-1}{k}+\binom{m-1}{k}\binom{n-1}{k-1}}{\binom{m+n}{n}}$$

自测题与练习

1. 某自助餐厅提供 3 道菜，人们可选择一道主菜、一种淀粉质食物和一份甜点心. 可能的选择如下表所示.

食品名称	选择
主菜	烤鸡、烤牛排
淀粉质食物	意大利面条、米饭、土豆
甜点心	冰淇淋、果冻、苹果馅饼、桃子

某人从每个种类中选择一道食物.

(a) 样本空间中有多少个结果?

(b) 令 A 表示"选择冰淇淋"的事件，A 包含多少个结果?

(c) 令 B 表示"选择烤鸡"的事件，B 包含多少个结果?

(d) 列出事件 AB 的所有可能结果.

(e) 令 C 表示"选择米饭"的事件，C 包含多少个结果?

(f) 列出事件 ABC 的所有结果.

61

2. 某顾客来到某商店的服装柜台，他以概率 0.22 购买套装，以概率 0.30 购买衬衣，以概率 0.28 购买领带，以概率 0.11 购买套装和衬衣，以概率 0.14 购买领带和套装，以概率 0.14 购买衬衣和领带，以概率 0.06 购买三种物品. 计算下列事件的概率:

(a) 三种物品都不购买.

(b) 恰好购买一种物品.

3. 分发一副扑克牌，第 14 张牌分到 A 的概率是多少? 第一张 A 出现在第 14 张牌的概率是多少?

4. 令 A 表示"洛杉矶市中心的温度为 70°F"的事件，B 表示"纽约市中心的温度为 70°F"事件，C 表示"纽约市中心和洛杉矶市中心的最高温度均为 70°F"的事件. 假如 $P(A)=0.3$，$P(B)=0.4$ 和 $P(C)=0.2$，求这两个城市的最低温度是 70°F 的概率.

5. 分发一副 52 张的普通扑克牌，计算下列事件的概率:

(a) 头 4 张牌的点数不同.

(b) 头 4 张牌花色不同的.

6. A 罐中有 3 个红球和 3 个黑球，B 罐中有 4 个红球和 6 个黑球. 若从每个罐中随机地抽取一个球，两个球颜色相同的概率是多少?

7. 在某彩票中，购买者在数字 1～40 中选择 8 个数字，然后彩票发行公司进行一次抽奖试验，从 40 个数字中摇出 8 个数字. 假定彩票发行公司是从 $\binom{40}{8}$ 种组合中等可能地选择一种组合，计算下列事件的概率:

(a) 彩票购买者 8 个数字全选对.

(b) 选对其中的 7 个数字.

(c) 至少选对其中的 6 个数字.

8. 在 3 名新生、4 名二年级学生、4 名三年级学生和 3 名四年级学生的人群中随机地选择 4 名学生组成一

个委员会，计算下列事件的概率：

(a) 该委员会由每个年级的一名学生组成.

(b) 该委员会由 2 名二年级学生和 2 名三年级学生组成.

(c) 该委员会仅由二年级或三年级学生组成.

9. 对一个有限集 A，令 $N(A)$表示 A 中元素的个数.

(a) 证明

$$N(A\cup B)=N(A)+N(B)-N(AB)$$

(b) 更一般地，证明

$$N\left(\bigcup_{i=1}^{n} A_i\right)=\sum_i N(A_i)-\sum_{i<j}\sum N(A_iA_j)+\cdots+(-1)^{n+1}N(A_1A_2\cdots A_n)$$

10. 考虑由标号为 1～6 的 6 匹马组成的赛马比赛的试验，假定样本空间由结束比赛的 6! 种可能排序组成. 令 A 表示“1 号马是前 3 名之一”的事件，B 表示“2 号马第二个到达终点”的事件，问事件 $A\cup B$ 有多少个结果？

11. 从完全洗好的 52 张牌中分发 5 张牌的一手牌，问这手牌中至少包含了四个花色的每个花色中的一张牌的概率是多少？

62

12. 一个篮球队由 6 个前锋和 4 个后卫组成，若给这些队员随机地分配宿舍，问正好有 2 个宿舍的队员是由一个后卫和一个前锋组成的概率是多少？

13. 假定某人从 R E S E R V E 中随机选择一个字母，然后再从 V E R T I C A L 中随机选择一个字母，问所选择的字母相同的概率是多少？

14. 证明布尔不等式：

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)\leqslant\sum_{i=1}^{\infty}P(A_i)$$

15. 证明：若 $P(A_i)=1$，对所有 $i\geqslant 1$，则 $P\left(\bigcap_{i=1}^{\infty} A_i\right)=1$.

16. 令 $T_k(n)$表示将集合$\{1,2,\cdots,n\}$分为 k 个非空子集的分割数，这里 $1\leqslant k\leqslant n$(参见理论练习 8 关于分割的定义). 论证：

$$T_k(n)=kT_k(n-1)+T_{k-1}(n-1)$$

提示：在一个子集中有多少个分割是$\{1\}$，在包含其他元素的子集中有多少个是 1 的部分？

17. 从包含 5 个红球、6 个白球和 7 个蓝球的罐中不放回地随机抽取 5 个球，求每个颜色的球至少有一个的概率.

63

第3章　条件概率与独立性

3.1　引言

本章介绍概率论中最重要的概念之一——条件概率．这个概念的重要性表现在两个方面：首先，在计算概率时，可以利用有关试验结果的部分信息，这时所求的概率就是条件概率．其次，即使没有部分信息可利用，条件概率作为一种工具，也可以比较容易地算出所求的概率．

3.2　条件概率

掷两个骰子，假定 36 种可能结果的发生是等可能的，即它们的概率都是 1/36．若已知第一个骰子是 3 点，那么在给出这一信息之后，两个骰子点数之和为 8 的概率是多少？为计算这一概率，进行如下推理：已知第一个骰子点数是 3，那么试验至多有 6 种可能结果，即(3,1)，(3,2)，(3,3)，(3,4)，(3,5)，(3,6)．这些结果的发生原来就有相同的概率，所以在已知第一个骰子点数是 3 的条件下，它们仍有相同的概率．即已知第一个骰子的点数是 3，结果(3,1)，(3,2)，(3,3)，(3,4)，(3,5)，(3,6)中每一个发生的概率是 1/6，在样本空间中其余 30 种结果的概率均为 0，因此所求的概率是 1/6．

令 E 与 F 分别表示两个骰子点数之和为 8 与第一个骰子点数是 3 这两个事件，那么前面算出的概率称为在 F 已经发生的条件下 E 发生的条件概率，记作 $P(E|F)$．

对任意事件 E 与 F 都适用的 $P(E|F)$ 的一般公式可以用同样的方法导出：若事件 F 已发生，则为使 E 也发生，必须是试验结果既在 E 中又在 F 中的某一点，即此点必属于 EF．由于知道 F 已发生，故 F 变成了新的或缩小的样本空间，于是 EF 发生的概率等于 EF 的概率与 F 的概率之比，一般地，可得到如下定义．

定义　若 $P(F)>0$，则

$$P(E|F)=\frac{P(EF)}{P(F)} \tag{2-1}$$

例 2a　将一枚硬币抛两次，假定样本空间 $S=\{(H,H),(H,T),(T,H),(T,T)\}$ 中的所有四个点是等可能的，已知第一次抛出正面，问两次都抛出正面的条件概率是多少？

解　若以 $E=\{(H,H)\}$ 表示两次都抛出正面的事件，而以 $F=\{(H,H),(H,T)\}$ 代表第一次抛出正面的事件，则所求的概率为

$$P(E|F)=\frac{P(EF)}{P(F)}=\frac{P(\{(H,H)\})}{P(\{(H,H),(H,T)\})}=\frac{\frac{1}{4}}{\frac{2}{4}}=\frac{1}{2}$$

■

例 2b　一个箱子里有 25 个灯泡，其中，5 个好灯泡最少能工作 30 天；10 个快坏了，第二天就不能用了；10 个完全坏了(不亮了)．随机地选择一个灯泡最初能用，求一星期后仍能用的概率．

解　令 G 表示随机选的灯泡为好灯泡的事件，T 表示灯泡不能用的事件．根据式(2-1)有

$$P(G|T^c)=\frac{P(GT^c)}{P(T^c)}$$

因为只有当所选的灯泡为好灯泡时，它才既是好灯泡又不是快坏的灯泡，所以 $GT^c=G$. 假定从25个灯泡中随机选取一个灯泡是等可能的，我们得到

$$P(G|T^c)=\frac{5/25}{15/25}=1/3$$

65

值得注意的是，上述概率可以通过缩小样本空间的方式来得到，因为我们已知所选择的灯泡不是完全坏的，问题就归结为从有10个不完全坏的灯泡和5个好灯泡的箱中随机地选一个好灯泡的概率. ■

若假定所有的结果是等可能的，通过缩小样本空间比直接利用式(2-1)更容易计算条件概率.

例2c 在桥牌比赛中，52张牌平均分发给东、南、西、北四个玩家，若南、北有8张黑桃，则东有剩余5张黑桃中的3张黑桃的概率是多少?

解 通过缩小样本空间很容易算出所求的概率. 已知南、北的26张牌中有8张黑桃，在剩余的26张牌中正好有5张黑桃在东、西的手中. 由于每种分发是等可能的，所以东有3张黑桃的概率为

$$\frac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}}\approx 0.339$$

■

例2d 琼斯的工作单位为至少有一个儿子的员工举行宴会，若已知琼斯有两个孩子，并且已应邀赴宴，问她的两个孩子都是男孩的条件概率是多少? 假定样本空间 $S=\{(b,b),(b,g),(g,b),(g,g)\}$，且所有的结果是等可能的. [这里 (b,g) 表示大的是男孩，而小的是女孩，其余类同.]

解 已知琼斯应邀赴宴与已知她至少有一个儿子是等可能的，这样，设 B 表示她的两个孩子均为男孩的事件，而 A 表示至少有一个孩子是男孩的事件，则所求的概率为

$$P(B|A)=\frac{P(BA)}{P(A)}=\frac{P(\{(b,b)\})}{P(\{(b,b),(b,g),(g,b)\})}=\frac{1/4}{3/4}=\frac{1}{3}$$

许多读者误以为在已知至少有一个孩子是男孩的条件下两个孩子均为男孩的条件概率为 $\frac{1}{2}$，而不是 $\frac{1}{3}$，因为他们认为不参加宴会的琼斯的孩子是一个男孩或是一个女孩是等可能的. 然而他们错在认为那两个概率是等可能的，因为最初四个结果是等可能的，已知至少有一个孩子是男孩等价于知道没有结果 (g,g). 于是我们只有 (b,b)，(b,g)，(g,b) 三种等可能的结果，因此琼斯的孩子中女孩不参加宴会的概率是男孩的两倍. ■

66

式(2-1)两边同乘 $P(F)$，得

$$P(EF)=P(F)P(E|F) \tag{2-2}$$

换言之，式(2-2)表示 E 与 F 同时发生的概率等于 F 发生的概率与给定 F 的条件下，E 发生的条件概率的积. 利用式(2-2)很容易计算事件交集的概率.

例2e 为选修法语课还是化学课这件事，Celine 犹豫不决. 她估计自己的法语课获A的概率是 $\frac{1}{2}$，而化学课是 $\frac{2}{3}$. 如果 Celine 通过抛一枚均匀硬币来决定此事，问她选修化学课且获A的概率是多少?

解 令 C 表示 Celine 选修化学课这一事件，而 A 表示她在所选的课程中获 A 的事件，则所求的概率为 $P(CA)$，利用等式(2-2)计算如下：

$$P(CA)=P(C)P(A|C)=\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)=\frac{1}{3}$$

■

例 2f 箱中有 8 个红球和 4 个白球，现不放回地取出 2 个球. 假定每次抽取时，箱中各个球被取出是等可能的，问取出的两个球都是红球的概率是多少？

解 令 R_1 和 R_2 分别为第一次与第二次抽取红球的事件. 若第一次取出的球是红球，则箱中还剩下 7 个红球和 4 个白球，从而 $P(R_2|R_1)=\frac{7}{11}$，而 $P(R_1)$ 显然等于 $\frac{8}{12}$，故所求的概率为

$$P(R_1R_2)=P(R_1)P(R_2|R_1)=\left(\frac{2}{3}\right)\left(\frac{7}{11}\right)=\frac{14}{33}$$

当然，此概率也可由

$$P(R_1R_2)=\frac{\binom{8}{2}}{\binom{12}{2}}$$

算出. ■

67

式(2-2)可推广到计算任意多个事件之交的概率，有时称为乘法公式.

乘法公式

$$P(E_1E_2E_3\cdots E_n)=P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1E_2\cdots E_{n-1})$$

对上式右边应用条件概率公式即可得到乘法公式. 即

$$P(E_1)\frac{P(E_1E_2)}{P(E_1)}\frac{P(E_1E_2E_3)}{P(E_1E_2)}\cdots\frac{P(E_1E_2\cdots E_n)}{P(E_1E_2\cdots E_{n-1})}=P(E_1E_2\cdots E_n)$$

我们利用乘法公式可以得到第 2 章例 5h(b)的另一种解法.

例 2g 52 张扑克牌随机地分成四份，每份 13 张. 计算每份中均有 1 个 A 的概率.

解 定义事件 $E_i(i=1,2,3,4)$ 如下：

$E_1=\{$黑桃 A 在任意一份中$\}$

$E_2=\{$黑桃 A 和红桃 A 在不同的份中$\}$

$E_3=\{$黑桃 A、红桃 A 和方块 A 在不同的份中$\}$

$E_4=\{$所有四个 A 在不同的份中$\}$

所求的概率为 $P(E_1E_2E_3E_4)$，利用乘法公式得

$$P(E_1E_2E_3E_4)=P(E_1)P(E_2|E_1)P(E_3|E_1E_2)P(E_4|E_1E_2E_3)$$

已知 $P(E_1)=1$，因此 E_1 是样本空间 S.

$$P(E_2|E_1)=\frac{39}{51}$$

因为包含黑桃 A 的那份中有剩余的 51 张牌中的 12 张，故

$$P(E_3|E_1E_2)=\frac{26}{50}$$

因为包含黑桃 A 和红桃 A 的那两份中分别有剩余 50 张牌中的 24 张，因此

$$P(E_4|E_1E_2E_3)=\frac{13}{49}$$

因此，我们得到每份中正好有一个 A 的概率为

68

$$P(E_1E_2E_3E_4)=\frac{39\cdot 26\cdot 13}{51\cdot 50\cdot 49}\approx 0.105$$

即每份中有一个 A 的概率近似为 10.5%.（习题 19 给出运用乘法公式解该问题的另一种解法.）■

注 我们关于 $P(E|F)$ 的定义与将概率作为最后的相对频率的解释是一致的. 为了说明这一点，假设作 n 次重复试验，其中 n 很大. 假定仅在 F 已经出现的情况下考虑这些试验，那么可以断定 $P(E|F)$ 将等于长期试验中 E 出现的比例. 为了验证这一点，注意到 $P(F)$ 是长期试验中 F 出现的比例，由此推出在 n 次重复试验中 F 近似出现 $nP(F)$ 次，类似地 E 和 F 同时出现 $nP(EF)$ 次. 因此，在 F 近似出现的 $nP(F)$ 次试验中，E 近似出现的比例等于

$$\frac{nP(EF)}{nP(F)}=\frac{P(EF)}{P(F)}$$

当 n 变得越来越大时，这个近似值越来越精确，这样我们得到了 $P(E|F)$ 的近似定义.

3.3 贝叶斯公式

设 E 和 F 是两个事件，可将 E 表示为 $E=EF\cup EF^c$. 这是因为，为使一个结果属于 E，它必须或者同属于 E 与 F，或者属于 E 而不属于 F（见图 3-1）. 由于 EF 和 EF^c 是互不相容的，由公理 3 有

$$\begin{aligned}P(E)&=P(EF)+P(EF^c)\\&=P(E|F)P(F)+P(E|F^c)P(F^c)\\&=P(E|F)P(F)+P(E|F^c)[1-P(F)]\end{aligned}\tag{3-1}$$

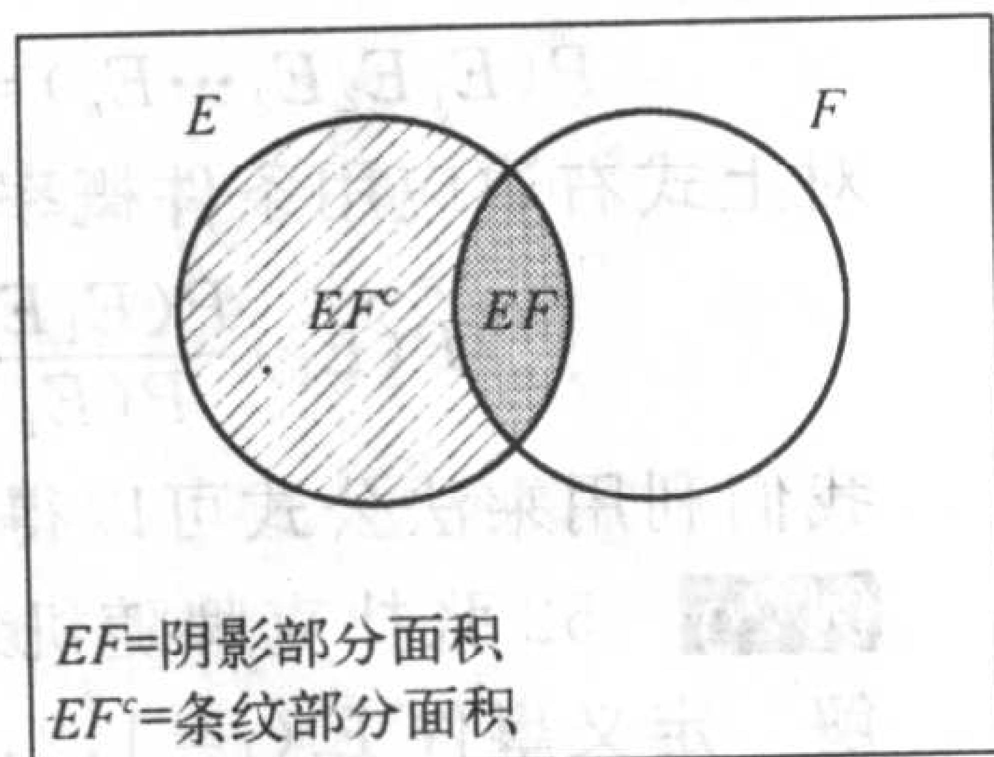


图 3-1　$E=EF\cup EF^c$

式(3-1)说明事件 E 的概率等于在 F 发生的条件下 E 的条件概率与在 F 不发生的条件下 E 的条件概率的加权平均，其中加在每个条件概率上的权正是作为条件的事件发生的概率. 这是一个极其有用的公式，因为它常常被用来利用第二个事件是否发生作为“条件”来确定某个事件的概率. 有许多例子很难直接计算某个事件的概率，但是，一旦我们知道了第二个事件是否发生，就可直接计算其概率. 下面举例来说明.

69

例 3a（第一部分） 某保险公司认为，人可以分为两类，一类是容易出事故的，另一类则比较谨慎. 他们的统计表明，一个易出事故的人在固定的一年内的某个时候出一次事故的概率是 0.4，而对于比较谨慎的人来说这个概率是 0.2. 若假定第一类人占 30%，那么，一个新保险客户在他购买保险后一年内将出一次事故的概率是多少？

解 以该保险客户是否是易出事故的人为条件，可以得到所求的概率. 令 A_1 为保险客户在一年期间出一次事故的事件，而以 A 表示保险客户是易出事故的人的事件，那么所求的概率 $P(A_1)$ 为

$$P(A_1)=P(A_1|A)P(A)+P(A_1|A^c)P(A^c)=(0.4)(0.3)+(0.2)(0.7)=0.26$$ ■

例 3a（第二部分） 如果一个新保险客户在他购买保险后一年内出了一次事故，问他是易出事故的人的概率是多少？

解 所求的概率 $P(A|A_1)$ 为

$$P(A|A_1)=\frac{P(AA_1)}{P(A_1)}=\frac{P(A)P(A_1|A)}{P(A_1)}=\frac{(0.3)(0.4)}{(0.26)}=\frac{6}{13}$$ ■

例 3b 考虑玩一副 52 张扑克牌的游戏. 洗好牌然后每次翻一张，在任何时候，玩家只要猜出下一张牌是黑桃 A 就算赢. 另外，若只剩下一张牌，黑桃 A 也没有出现且前几次玩家均没有猜对，也可以认为玩家赢. 什么是好策略，什么是坏策略?

解 每种策略赢的概率为 $\frac{1}{52}$. 运用数学归纳法可以证明更强的结论，不管采用哪种策略，从 n 张牌中猜中黑桃 A，赢的概率为 $\frac{1}{n}$. 当 $n=1$ 时，结论显然成立. 假定对 $n-1$ 张牌结论仍然成立. 对 n 张牌，固定任何一种策略，令 P 表示这种策略猜对第一张牌是黑桃 A 的概率，则玩家赢的概率为 $\frac{1}{n}$. 另一方面，若这种策略没有猜对第一张牌是黑桃 A，则玩家赢的概率是第一张不是黑桃 A 的概率(即 $\frac{n-1}{n}$)乘以给定第一张不是黑桃 A 玩家赢的条件概率. 但后一个条件概率等于 $n-1$ 张牌中只有一个黑桃 A 玩家赢的概率，由归纳法可知条件概率为 $\frac{1}{n-1}$. 因此采用没有猜对第一张是黑桃 A 的策略，玩家赢的概率为

70

$$\frac{n-1}{n}\frac{1}{n-1}=\frac{1}{n}$$

所以，令 G 表示猜对第一张牌的事件，我们可以得到

$$P\{赢\}=P\{赢|G\}P(G)+P\{赢|G^c\}(1-P(G))=\frac{1}{n}p+\frac{1}{n}(1-p)=\frac{1}{n}$$ ■

例 3c 在多项选择题中，某考生可能知道正确答案，也可能是乱猜一个. 令他知道正确答案的概率是 p，而乱猜的概率是 $1-p$. 设他乱猜答案猜对的概率是 $\frac{1}{m}$，这里 m 是多项选择的个数. 如果已知他答对了，问他确实知道考题正确答案的条件概率是多少?

解 令 C 与 K 分别表示这个考生答对了与他知道正确答案这两个事件，则有

$$P(K|C)=\frac{P(KC)}{P(C)}=\frac{P(C|K)P(K)}{P(C|K)P(K)+P(C|K^c)P(K^c)}$$

$$=\frac{p}{p+(1/m)(1-p)}=\frac{mp}{1+(m-1)p}$$

71

例如，若 $m=5$，$p=\frac{1}{2}$，则此概率是 $\frac{5}{6}$. ■

例 3d 一项血液化验有 95%的把握将患某种疾病的人鉴别出来. 但是，这项化验用于健康人也会有 1%的“伪阳性”结果.（这就是说，如果一个健康人接受这项化验，则化验结果误诊此人患这种疾病的概率是 0.01. ）如果这种疾病的患者事实上仅占人口的 0.5%，而某人化验结果是阳性，问此人确实患这种疾病的概率是多少?

解 令 D 表示接受化验的人患有这种疾病的事件，E 表示他的化验结果为阳性这一事件，所求的概率 $P(D|E)$ 为

$$P(D|E)=\frac{P(DE)}{P(E)}=\frac{P(E|D)P(D)}{P(E|D)P(D)+P(E|D^c)P(D^c)}$$

$$=\frac{(0.95)(0.005)}{(0.95)(0.005)+(0.01)(0.995)}=\frac{95}{294}\approx 0.323$$

因此，在验血结果为阳性的人当中，真正患这种疾病的只占 32%，对于这一结果，许多学生感到很吃惊(因为验血似乎是个好办法，他们期望这个数值应该高得多). 因此提出第二种解法是值得的，与前一种解法比较，第二种解法尽管不怎么严格，但可能更直观.

由于事实上患此病的人占人口的 0.5%，平均来看，接受化验的每 200 个人中应有 1 个患者. 而这项化验只能保证此人确实患此病的概率是 0.95. 因此，可以认为每 200 个人中能保证有 0.95 个人确实患此病. 但另一方面，在其余 199 个健康人中，这项化验会错误地诊断出 199×0.01个人患此病. 因此每诊断出 0.95 个病人时，总有 199×0.01 个健康人误诊为患病. 于是，当验血结果判定某人患此病时，正确的诊断所占的比例为

$$\frac{0.95}{0.95+(199)(0.01)}=\frac{95}{294}\approx 0.323$$

■

72

式(3-1)对于根据附加信息求个人概率仍是有用的. 见下例.

例 3c 某医生陷入一种困境:“如果我至少有 80%的可能确信我的患者有这种病，一般会建议患者做手术，而如果我不太确信，则会建议患者再做一些检查，这些检查可能是昂贵的，也可能有些疼痛. 现在，最初我只有 60%的可能确定琼斯有这种病，于是让他做检查 A，若患者有这种病或他不是很健康，检查 A 会给出一个阳性结果. 检查结果是阳性，琼斯告诉我他有糖尿病，于是我建议他做手术. 不管前面的信息多么复杂，不会改变我最初有 60%的可能确定他有这种病，但它会影响检查 A 的结论. 这是因为当患者是健康的而不会产生阳性结果时，在糖尿病患者不会患这种病的情形下，检查 A 还是有 30%的可能给出阳性检查结果. 现在我应该怎么办? 做更多的检查还是立即手术?”

解 为了决定是否要做手术，医生首先应该计算出检查 A 的结果是阳性时，琼斯有这种病的修正概率. 令 D 表示琼斯有这种病的事件，E 表示检查 A 的结果是阳性的事件，则所求的条件概率 $P(D|E)$为

$$P(D|E)=\frac{P(DE)}{P(E)}=\frac{P(D)P(E|D)}{P(E|D)P(D)+P(E|D^c)P(D^c)}$$

$$=\frac{(0.6)1}{1(0.6)+(0.3)(0.4)}=0.833$$

利用琼斯是否有这种病作为条件，然后利用因为琼斯有糖尿病所以他没有这种病，而检查结果呈阳性的条件概率 $P(E|D^c)$为 0.3 这一事实，我们可以计算出检查结果为阳性的概率. 此时医生应该至少有 80%的可能确信琼斯有这种病，这时他就可以建议琼斯做手术. ■

例 3f 根据犯罪记录可知确定的嫌疑犯中有 60%是有罪的. 假定现在有一些新的证据说明嫌疑犯有一些特征(如左撇子、秃头或棕色头发)，若 20%的人有这些特征，问如果查出嫌疑犯有这些特征，则证明他有罪的概率是多少?

73

解 令 G 表示嫌疑犯有罪的事件，C 表示他有这些特征的事件，我们得到

$$P(G|C)=\frac{P(GC)}{P(C)}=\frac{P(C|G)P(G)}{P(C|G)P(G)+P(C|G^c)P(G^c)}$$

$$=\frac{1(0.6)}{1(0.6)+(0.2)(0.4)}\approx 0.882$$

其中我们假定嫌疑犯是无罪的但有这些特征的概率为 0.2. ■

例 3g 1965 年 5 月在布宜诺斯艾利斯(阿根廷首都)举办的世界桥牌锦标赛中，著名的英国桥牌搭档 Terrence Reese 和 Boris Schapiro 被控告利用手指传递信号的方式作弊，这种方式能够暗示参赛者持有多少张红桃牌. Reese 和 Schapiro 否认指控，最终由英国桥牌协会举行听证会. 听证会按诉讼程序进行，原告和被告两方都有权传唤和相互指认目击者. 在听证过程中，原告检查了 Reese 和 Schapiro 打的几手特定的牌，声称他们打的这几手牌与他们不正当地获知红桃牌的违法前提一致. 关于这一点，被告律师指出他们打的这几手牌和他们的规范打法完全一致. 然而，原告声称只要他们的打法和不正当的前提相符，则它就可以作为支持这个前提的证据. 你如何评判原告的推理?

解 问题的关键之一是新证据(在本例中，指的是几手牌的打法)如何影响这个前提的概率. 现在令 H 表示这个前提(例如，Reese 和 Schapiro 不正当地获知红桃牌)，E 表示新证据，则

$$P(H|E)=\frac{P(HE)}{P(E)}=\frac{P(E|H)P(H)}{P(E|H)P(H)+P(E|H^c)[1-P(H)]} \tag{3-2}$$

其中 $P(H)$ 表示引入新证据前的似然估计，只要新证据使这个前提更有可能，即 $P(H|E)\geqslant P(H)$，则新证据将支持这个前提. 从式(3-2)得到

74

$$P(E|H)\geqslant P(E|H)P(H)+P(E|H^c)[1-P(H)]$$

或等价于

$$P(E|H)\geqslant P(E|H^c)$$

换言之，任何新证据被认为支持这个前提，仅当新证据更可能使这个前提成立而不是更可能使这个前提不成立. 事实上，这个前提的新概率依赖于它的最初概率以及这些条件概率的比，从式(3-2)得到

$$P(H|E)=\frac{P(H)}{P(H)+[1-P(H)]\dfrac{P(E|H^c)}{P(E|H)}}$$

因此，考虑这个问题，牌的打法被认为与不正当的前提相符，仅当这样的打法更可能是他们在作弊. 因为原告没有做出这一声明，所以他认为与不正当的前提相符的证据是无效的. ■

当引入新证据时，前提的概率的变化可以根据这个前提的优势率(odds ratio)的变化来简洁地表示，关于优势率的概念可以如下定义.

定义 事件 A 的优势率为

$$\frac{P(A)}{P(A^c)}=\frac{P(A)}{1-P(A)}$$

也就是说，事件 A 的优势率说明事件 A 发生的概率在多大程度上更可能大于事件 A 不发生的概率，例如，若 $P(A)=\frac{2}{3}$，则 $P(A)=2P(A^c)$，于是优势率为 2. 若优势率等于 α，则利用前提可以说优势率"$\alpha\to1$".

现在考虑前提 H 以概率 $P(H)$ 是成立的，假定引入一个新的证据 E，给定 E 的条件下，则 H 成立与 H 不成立的条件概率为

$$P(H|E)=\frac{P(E|H)P(H)}{P(E)} \qquad P(H^c|E)=\frac{P(E|H^c)P(H^c)}{P(E)}$$

因此，引入证据 E 后的新优势率为

75

$$\boxed{\frac{P(H|E)}{P(H^c|E)}=\frac{P(H)}{P(H^c)}\frac{P(E|H)}{P(E|H^c)}} \tag{3-3}$$

也就是说，H 的新优势率为其旧优势率乘以新证据使 H 成立的条件下的条件概率与新证据使 H 不成立的条件下的条件概率之比．这证实了例 3f 的结果，因为只要新证据更可能使 H 成立而不是更可能使 H 不成立，优势率以及 H 的概率增大．同样，只要新证据更可能使 H 不成立而不是更可能使 H 成立，优势率减小．

例 3h 抛硬币 A 出现正面的概率为$\frac{1}{4}$，而抛硬币 B 出现正面的概率为$\frac{3}{4}$．现随机地选定一枚硬币抛两次，如果两次均出现正面，问抛的是硬币 B 的概率是多少？

解 令 B 为抛的是硬币 B 的事件，由于 $P(B)=P(B^c)$，由式(3-3)得

$$\frac{P(B|\text{两次正面})}{P(B^c|\text{两次正面})}=\frac{9/16}{1/16}=9$$

所以优势率是 9，或等价地，抛的是硬币 B 的概率为$\frac{9}{10}$． ■

式(3-1)可以作如下推广：设 F_1，F_2，…，F_n 是互不相容的事件，且满足

$$\bigcup_{i=1}^{n} F_i = S$$

换言之，一次试验的结果正好是事件 F_1，F_2，…，F_n 中的某一个发生．注意到

$$E=\bigcup_{i=1}^{n} EF_i$$

而且事件 $EF_i(i=1,2,\cdots,n)$是互不相容的，我们得到

$$P(E)=\sum_{i=1}^{n}P(EF_i)=\sum_{i=1}^{n}P(E|F_i)P(F_i) \tag{3-4}$$

式(3-4)说明，如果 F_i 中的一个事件发生(因为事件 F_1，F_2，…，F_n 中有且仅有一个必发生)，就可以以该事件作为条件计算出 $P(E)$．也就是说，式(3-4)表明 $P(E)$是 $P(E|F_i)$的加权平均，其中每一项是通过作为条件的事件的概率进行加权的．

76

假定 E 发生，下面确定计算 F_j 发生的概率．由式(3-4)可以得到如下命题．

命题 3.1

$$P(F_j|E)=\frac{P(EF_j)}{P(E)}=\frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n}P(E|F_i)P(F_i)} \tag{3-5}$$

式(3-5)称为贝叶斯公式，来源于英国哲学家托马斯·贝叶斯(Thomas Bayes)．如果我们把这些事件 F_j 设想为关于某件事情的各个可能的“假设条件”，那么贝叶斯公式可以这样理解：在试验之前对这些假设条件成立所做的判断[即 $P(F_j)$]，应该如何根据试验的结果进行修改．

例 3i 一架飞机失事了，假定它在任意三个地区的任一个坠毁是等可能的．令 $1-\beta_i$ 表示在第 i 个地区搜索找到飞机的概率，其中 $i=1$，2，3．(常数 β_i 称为未发现概率(overlook probability)，表示没注意到飞机的概率，通常可归因于该地区的地理和环境条件所致)．问在第一个地区搜索未找到飞机，而在第 $i(i=1,2,3)$个地区找到飞机的条件概率是多少？

解 令 $R_i(i=1,2,3)$表示飞机在第 i 个地区的事件，E 表示在第一个地区搜索未找到飞机的事件．由贝叶斯公式，我们得到

$$P(R_1|E)=\frac{P(ER_1)}{P(E)}=\frac{P(E|R_1)P(R_1)}{\sum_{i=1}^{3}P(E|R_i)P(R_i)}$$

$$=\frac{(\beta_1)\frac{1}{3}}{(\beta_1)\frac{1}{3}+(1)\frac{1}{3}+(1)\frac{1}{3}}=\frac{\beta_1}{\beta_1+2}$$

77

对于 $j=2, 3$,

$$P(R_j|E)=\frac{P(E|R_j)P(R_j)}{P(E)}=\frac{(1)\frac{1}{3}}{(\beta_1)\frac{1}{3}+\frac{1}{3}+\frac{1}{3}}=\frac{1}{\beta_1+2}\qquad j=2, 3$$

值得注意的是，给定了在第一个地区搜索未找到飞机的先验信息，飞机在第 j 个地区的修正概率(即条件概率)大于 $j\neq 1$ 时飞机在第 j 个地区的最初概率，而小于 $j=1$ 时的最初概率. 从直观上来讲，在第一个地区搜索未找到飞机应该减少了它在该地区的可能而增加了在其他地区的可能. 同样，飞机在第一个地区而没有被找到的条件概率是未发现概率 β_1 的增函数，因为 β_1 越大，越有可能没注意到飞机，与飞机不在那里正相反. 同样地，当 $j\neq 1$ 时，$P(R_j|E)$是 β_1 的减函数. ■

下一个例子经常被学过概率而又不讲道德的学生用来从他们的知识较少的朋友那里赢钱.

例 3j 假设有三张形状完全相同但所涂颜色不同的卡片，第一张两面全是红色的，第二张两面全是黑色的，而第三张是一面红一面黑. 现将三张卡片放在一个帽子里充分混合后，随机取出一张卡片放在地面上，若选出的卡片朝上的一面是红色的，那么另一面是黑色的概率是多少?

解 令 RR、BB 与 RB 分别表示取出的卡片是两面红、两面黑与一面红一面黑这三个事件. 令 R 表示选出的卡片朝上的一面是红色这一事件，因此所求的概率为

$$P(RB|R)=\frac{P(RB\cap R)}{P(R)}=\frac{P(R|RB)P(RB)}{P(R|RR)P(RR)+P(R|RB)P(RB)+P(R|BB)P(BB)}$$

$$=\frac{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)}{(1)\left(\frac{1}{3}\right)+\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)+0\left(\frac{1}{3}\right)}=\frac{1}{3}$$

于是答案是$\frac{1}{3}$，很多学生误以为出现一面是红色，答案是$\frac{1}{2}$. 他们的理由是，已经出现了红色的一面，那么就有两种等可能的结果：这张卡片的两面都是红的，或者一面红一面黑. 但是他们的错误正是假定上述两种结果是等可能的. 这是因为，如果认为每张卡片由两个不同的面组成，那么试验就有 6 个等可能的结果，即 R_1, R_2, B_1, B_2, R_3, B_3，其中 R_1 表示红卡片的第一个红面朝上，R_2 表示红卡片的第二个红面朝上，R_3 表示红黑卡片的红面朝上，等等. 因为仅当朝上的红面是 R_3 时，卡片的另一面才是黑的，故所求的概率是在“R_1, R_2, R_3 中某一个发生”的条件下的 R_3 的条件概率，显然应等于$\frac{1}{3}$. ■

78

例 3k 已知有两个孩子的一对新夫妇刚搬到某小城镇. 假定有人在路上遇见母亲与她的一个孩子散步，若这个孩子是女孩，问她的两个孩子都是女孩的概率是多少?

解 定义如下事件：

G_1：第一个孩子(即年长)是女孩.

G_2：第二个孩子是女孩.

G：与母亲在一起的孩子是女孩.

又令 B_1，B_2，B 表示由“男孩”代替“女孩”的类似事件，所求的概率 $P(G_1G_2|G)$为

$$P(G_1G_2|G)=\frac{P(G_1G_2G)}{P(G)}=\frac{P(G_1G_2)}{P(G)}$$

又

$$\begin{aligned}P(G)&=P(G|G_1G_2)P(G_1G_2)+P(G|G_1B_2)P(G_1B_2)+P(G|B_1G_2)P(B_1G_2)\\&\quad+P(G|B_1B_2)P(B_1B_2)\\&=P(G_1G_2)+P(G|G_1B_2)P(G_1B_2)+P(G|B_1G_2)P(B_1G_2)\end{aligned}$$

上式第二个等号用到 $P(G|G_1G_2)=1$ 和 $P(G|B_1B_2)=0$. 若假设四种性别组合是等可能的，则

$$P(G_1G_2|G)=\frac{\frac{1}{4}}{\frac{1}{4}+\frac{P(G|G_1B_2)}{4}+\frac{P(G|B_1G_2)}{4}}=\frac{1}{1+P(G|G_1B_2)+P(G|B_1G_2)}$$

因此，答案依赖于给定事件 G_1B_2、与母亲在一起的孩子是女孩以及给定事件 G_2B_1、与母亲在一起的孩子是女孩的条件概率的假定. 例如，如果假定孩子的性别是独立的，与母亲在一起的孩子是年长的孩子的概率为 p，则有

79

$$P(G|G_1B_2)=p=1-P(G|B_1G_2)$$

由上式可以得到

$$P(G_1G_2|G)=\frac{1}{2}$$

另一方面，若假定这两个孩子性别不同，母亲选择和女孩一起散步的概率为 q，这两个孩子的出生顺序是独立的，则可以得到

$$P(G|G_1B_2)=P(G|B_1G_2)=q$$

由上式可以得到

$$P(G_1G_2|G)=\frac{1}{1+2q}$$

例如，若取 $q=1$，即母亲总选择和女儿一起散步，则两个女儿的条件概率为$\frac{1}{3}$，这和例 2d 一致，因为看到母亲现在与女儿在一起等价于至少有一个女孩的事件.

因此，如上所述，问题是不可解的. 的确，即使做了通常的性别概率的等可能性假设，在求解之前仍需作附加的假设. 这是因为试验的样本空间由形式为 s_1，s_2，i 的向量组成，其中 s_1 是年长孩子的性别，s_2 是年幼孩子的性别，i 表示与母亲在一起的孩子的出生次序. 因此为了确定样本空间的事件的概率，仅仅作关于孩子的性别的假设是不够的，还需对已知与母亲在一起的孩子的性别的条件下的条件概率作某些假设. ■

例 3l 箱中有可供使用的三种型号的手电筒，第一种型号的手电筒使用超过 100 小时的概率为 0.7，第二种型号的手电筒和第三种型号的手电筒的相应概率分别为 0.4 与 0.3. 假定箱中有 20%的第一种型号的手电筒、30%的第二种型号的手电筒、50%的第三种型号的手电筒.

(a) 随机取出一个手电筒使用超过 100 小时的概率为多少?

(b) 给定的手电筒使用超过 100 小时，则它是第 $j(j=1,2,3)$种型号的手电筒的条件概率

为多少？

解 (a) 令 A 表示取出一个手电筒使用超过 100 小时的事件，F_j 表示取出第 $j(j=1,2,3)$种型号的手电筒的事件．为计算 $P(A)$，以手电筒的型号为条件得到

$$\begin{aligned}P(A)&=P(A|F_1)P(F_1)+P(A|F_2)P(F_2)+P(A|F_3)P(F_3)\\&=(0.7)(0.2)+(0.4)(0.3)+(0.3)(0.5)=0.41\end{aligned}$$

80

手电筒使用超过 100 小时的概率为 41%．

(b) 利用贝叶斯公式得到所求的概率为

$$P(F_j|A)=\frac{P(AF_j)}{P(A)}=\frac{P(A|F_j)P(F_j)}{0.41}$$

于是，

$$P(F_1|A)=\frac{(0.7)(0.2)}{0.41}=\frac{14}{41}$$

$$P(F_2|A)=\frac{(0.4)(0.3)}{0.41}=\frac{12}{41}$$

$$P(F_3|A)=\frac{(0.3)(0.5)}{0.41}=\frac{15}{41}$$

例如，尽管箱中最初有第一种型号的手电筒的概率为 0.2，但使用超过 100 小时的概率上升为$\frac{14}{41}\approx 0.341$. ■

例 3m 一个罪犯单独作案，在犯罪现场留下他的 DNA 信息．研究 DNA 的法医们发现只有 5 条染色体是可以识别的，每个无罪的人有 5 条染色体匹配的 DNA 信息的概率为 10^{-5}．检察官认为罪犯可能是镇里 1 000 000 居民中的一个．10 000 人在近十年内出狱，因此有他们的 DNA 记录．在分析 DNA 信息之前，检察官认为有前科的 10 000 人这次犯罪的概率为 α，而剩余 990 000 人这次犯罪的概率为 β，其中 $\alpha=c\beta$(即假定近期出狱的人这次犯罪的可能性是其他居民这次犯罪的可能性的 c 倍)．分析有前科的 10 000 人的 DNA 信息，发现 AJ Jones 是唯一的 DNA 信息匹配的人．假设检察官对 α 和 β 的估计是正确的，问 AJ 有罪的概率是多少？

解 因为概率和为 1，我们可以得到

$$1=10\,000\alpha+990\,000\beta=(10\,000c+990\,000)\beta$$

于是，

$$\beta=\frac{1}{10\,000c+990\,000},\quad \alpha=\frac{c}{10\,000c+990\,000}$$

现令 G 表示 AJ 有罪的事件，M 表示 AJ 是有 DNA 记录的 10 000 人中唯一匹配的事件，则

81

$$P(G|M)=\frac{P(GM)}{P(M)}=\frac{P(G)P(M|G)}{P(M|G)P(G)+P(M|G^c)P(G^c)}$$

若 AJ 有罪，则他是有 DNA 记录的 10 000 人中唯一匹配的人，而其他人均不匹配．因此，

$$P(M|G)=(1-10^{-5})^{9\,999}$$

另一方面，若 AJ 无罪，他也是有 DNA 记录的 10 000 人中唯一匹配的人(发生的概率为 10^{-5})，所有其他有 DNA 记录的人均无罪并且不匹配．现假定 AJ 无罪，则所有其他有记录的人也无罪的条件概率为

$$P(\text{其他人无罪}|\text{AJ 无罪})=\frac{P(\text{所有有记录的人均无罪})}{P(\text{AJ 无罪})}=\frac{1-10\,000\alpha}{1-\alpha}$$

另外，假定他们均无罪，有记录的其他人有匹配的条件概率为

$$P(M|G^c)=10^{-5}\left(\frac{1-10\,000\alpha}{1-\alpha}\right)(1-10^{-5})^{9\,999}$$

因为 $P(G)=\alpha$，则由上式可得

$$P(G|M)=\frac{\alpha}{\alpha+10^{-5}(1-10\,000\alpha)}=\frac{1}{0.9+\frac{10^{-5}}{\alpha}}$$

若检察官最初认为有前科的人这次犯罪的概率为没有前科的人的 100 倍(即 $c=100$)，则 $\alpha=\frac{1}{19\,900}$，此时

$$P(G|M)=\frac{1}{1.099}\approx 0.909\,9$$

若检察官最初认为适当的比率 $c=10$，则 $\alpha=\frac{1}{109\,000}$，此时

82

$$P(G|M)=\frac{1}{1.99}\approx 0.502\,5$$

若检察官最初认为罪犯可能是镇里的任意一个人($c=1$)，则 $\alpha=10^{-6}$，此时

$$P(G|M)=\frac{1}{10.9}\approx 0.091\,7$$

若检察官最初认为所有的人有同样的可能犯罪，则所求的概率近似为 9%，若检察官认为有前科的人犯罪是没有前科的人的 100 倍，这时概率近似为 91%. ■

3.4 独立事件

本章前面的例子表明，在条件 F 下 E 的条件概率 $P(E|F)$ 一般不等于 E 的无条件概率 $P(E)$. 换言之，已知 F 已发生后一般要改变 E 发生的可能性. 在 $P(E|F)$ 与 $P(E)$ 相等的特殊情况下，称 E 独立于 F. 这就是说，若 F 已发生这一信息不改变 E 发生的概率，则 E 独立于 F.

由于 $P(E|F)=P(EF)/P(F)$，因此，如果

$$P(EF)=P(E)P(F) \tag{4-1}$$

则 E 独立于 F，式(4-1)中 E 与 F 是对称的，这表明若 E 独立于 F，则 F 也独立于 E. 于是，我们有如下定义.

定义 *如果式(4-1)成立，则两个事件 E 与 F 称为是独立的.*

不独立的两个事件 E 与 F 称为是相关的.

例 4a 从一副 52 张扑克牌中随机抽取一张，以 E 表示抽取的牌是 A 的事件，而 F 表示抽取的牌是黑桃的事件，则 E 与 F 是独立的. 这是因为 $P(EF)=\frac{1}{52}$，而 $P(E)=\frac{4}{52}$，$P(F)=\frac{13}{52}$. ■

例 4b 抛两枚硬币，设所有的四个结果是等可能的. 若 E 表示第一枚硬币出现正面的事件，而 F 表示第二枚硬币出现反面的事件，则 E 与 F 是独立的，因为 $P(EF)=P(\{(H,T)\})=\frac{1}{4}$；而 $P(E)=P(\{(H,H),(H,T)\})=\frac{1}{2}$，$P(F)=P(\{(H,T),(T,T)\})=\frac{1}{2}$. ■

83

例 4c 掷两个均匀的骰子，以 E_1 表示两个骰子点数之和为 6 的事件，F 表示第一个骰子点数是 4 的事件，则

$$P(E_1F)=P(\{(4,2)\})=\frac{1}{36}$$

而

$$P(E_1)P(F)=\left(\frac{5}{36}\right)\left(\frac{1}{6}\right)=\frac{5}{216}$$

所以 E_1 和 F 是不独立的．直观地看，原因是明显的．如果我们感兴趣的是两个骰子点数之和为 6 的可能性，那么当第一个骰子点数是 4(或点数为 1，2，3，4，5 中的任何一个)时，我们是相当高兴的，因为这时仍有可能得到总点数 6．从另一方面说，当第一个骰子为 6 点时，我们就不高兴了，因为不再有总点数为 6 的可能性．换言之，两个骰子点数之和为 6 的可能性依赖于掷第一个骰子所得的结果，因此 E_1 和 F 不独立．

现在，设 E_2 表示两个骰子点数之和为 7 这一事件，E_2 和 F 是否独立？答案是肯定的，因为

$$P(E_2F)=P(\{(4,3)\})=\frac{1}{36}$$

而

$$P(E_2)P(F)=\left(\frac{1}{6}\right)\left(\frac{1}{6}\right)=\frac{1}{36}$$

为什么两个骰子点数之和为 7 这个事件与第一个骰子掷出的结果是独立的呢？对此可以给出一个直观论证，我们将它留给读者． ■

例 4d 令 E 表示下一任总统是共和党人的事件，F 表示下一年出现大地震的事件，大多数人认为 E 与 F 是独立的．令 G 表示选举后两年内出现经济萧条的事件，关于假设 E 与 G 是独立的可能会出现争论． ■

我们现在说明若 E 与 F 独立，则 E 与 F^c 也独立．

命题 4.1 *若事件 E 与 F 独立，则 E 与 F^c 也独立．*

证明 假定 E 与 F 独立，由于 $E=EF\cup EF^c$，而 EF 和 EF^c 显然互不相容，所以有

$$P(E)=P(EF)+P(EF^c)=P(E)P(F)+P(EF^c)$$

84

或等价地，有

$$P(EF^c)=P(E)[1-P(F)]=P(E)P(F^c)$$

命题得证． ■

这样，若 E 与 F 独立，则有关 F 是否发生的信息不会改变 E 发生的概率．

现设事件 E 既与 F 独立，又与 G 独立，那么 E 是否必定与 FG 独立？令人稍感吃惊的是，答案是否定的．考虑下面的例子．

例 4e 掷两个均匀的骰子，令 E 表示两个骰子点数之和为 7 的事件，F 表示第一个骰子点数为 4 的事件，而 G 表示第二个骰子点数为 3 的事件．由例 4c 可知 E 与 F 独立，同理可知 E 与 G 也是独立的，很明显 E 和 FG 是不独立的．[因为 $P(E|FG)=1$．] ■

受例 4e 的启发，将在下面给出三个事件 E，F 与 G 独立的适当定义，这个定义比仅要求 $\binom{3}{2}$ 对事件独立更进一步，因此我们可得到如下定义．

定义 三个事件 E，F，G 称为是独立的，如果

$$P(EFG)=P(E)P(F)P(G)$$
$$P(EF)=P(E)P(F)$$
$$P(EG)=P(E)P(G)$$
$$P(FG)=P(F)P(G)$$

应该指出，若事件 E, F, G 独立，那么 E 与 F 和 G 组成的任何事件独立. 例如，E 与 $F\cup G$ 独立，这是因为

$$\begin{aligned}P[E(F\cup G)]&=P(EF\cup EG)=P(EF)+P(EG)-P(EFG)\\&=P(E)P(F)+P(E)P(G)-P(E)P(FG)\\&=P(E)[P(F)+P(G)-P(FG)]=P(E)P(F\cup G)\end{aligned}$$

85

当然，对于多于三个的一组事件，也可以类似地定义它们的独立性. 事件 E_1, E_2, …, E_n 称为是独立的，如果对这 n 个事件中的任意 r 个事件 $E_{1'}$, $E_{2'}$, …, $E_{r'}$ $(r\leqslant n)$ 有

$$P(E_{1'}E_{2'}\cdots E_{r'})=P(E_{1'})P(E_{2'})\cdots P(E_{r'})$$

最后，称无限多个事件是独立的，如果其中任意有限多个事件是独立的.

有时会遇到这种情况，所考虑的概率试验由一系列子试验组成. 例如，连续抛一枚硬币这个试验，就可以把每抛一次看成一个子试验. 在许多情形下，假定任一组子试验的结果不影响其他子试验的结果的概率是合理的. 如果是这样的情形，我们称这些子试验是独立的. 更正式地说，我们称一系列子试验是独立的，如果 E_1, E_2, …, E_n, …一定是独立的事件序列，其中，事件 E_i 的发生完全由第 i 次子试验的结果所决定.

如果各个子试验彼此相同，即各子试验有相同的(子)样本空间及相同的定义在其事件上的概率函数，那么就称这些子试验为**重复试验**(trials).

例 4f 进行一个独立的重复试验无穷序列，每次试验结果成功的概率为 p，失败的概率为 $1-p$，试求下列概率：

(a) 前 n 次试验中至少成功 1 次.

(b) 前 n 次试验中正好成功 k 次.

(c) 所有试验结果都成功.

解 为计算前 n 次试验中至少成功 1 次的概率，先求前 n 次试验全失败的概率. 若以 E_i 表示第 i 次试验失败这一事件，则由独立性可知前 n 次试验全失败的概率为

$$P(E_1E_2\cdots E_n)=P(E_1)P(E_2)\cdots P(E_n)=(1-p)^n$$

因此，(a)的答案为 $1-(1-p)^n$.

为解(b)，考虑任一个由 k 次成功、$n-k$ 次失败组成的前 n 个结果的特定序列，由独立性可知，每个这样的序列发生的概率均为 $p^k(1-p)^{n-k}$. 因为这样的序列总共有 $\binom{n}{k}$ 个(由 k 次成功、$n-k$ 次失败组成的排列数为 $n!/k!(n-k)!$)，故(b)所求的概率为

$$P\{恰好\ k\ 次成功\}=\binom{n}{k}p^k(1-p)^{n-k}$$

86

为解(c)，我们注意到，由(a)可知前 n 次试验全成功的概率为

$$P(E_1^cE_2^c\cdots E_n^c)=p^n$$

于是，运用概率的连续性(见 2.6 节)可得所求的概率 $P\left(\bigcap_{i=1}^{\infty}E_i^c\right)$ 为

$$\begin{aligned}P\left(\bigcap_{i=1}^{\infty}E_i^c\right)&=P\left(\lim_{n\to\infty}\bigcap_{i=1}^{n}E_i^c\right)=\lim_{n\to\infty}P\left(\bigcap_{i=1}^{n}E_i^c\right)\\&=\lim_n p^n=\begin{cases}0 & 若\ p<1\\1 & 若\ p=1\end{cases}\end{aligned}$$

■

例 4g 由 n 个元件组成的一个系统，若至少一个元件能工作，整个系统仍能工作(见图 3-2)，称为并联系统. 对于一个这样的系统，若元件 i 和其他元件独立，并以概率 $p_i(i=1,2,\cdots,n)$工作，问系统正常工作的概率是多少?

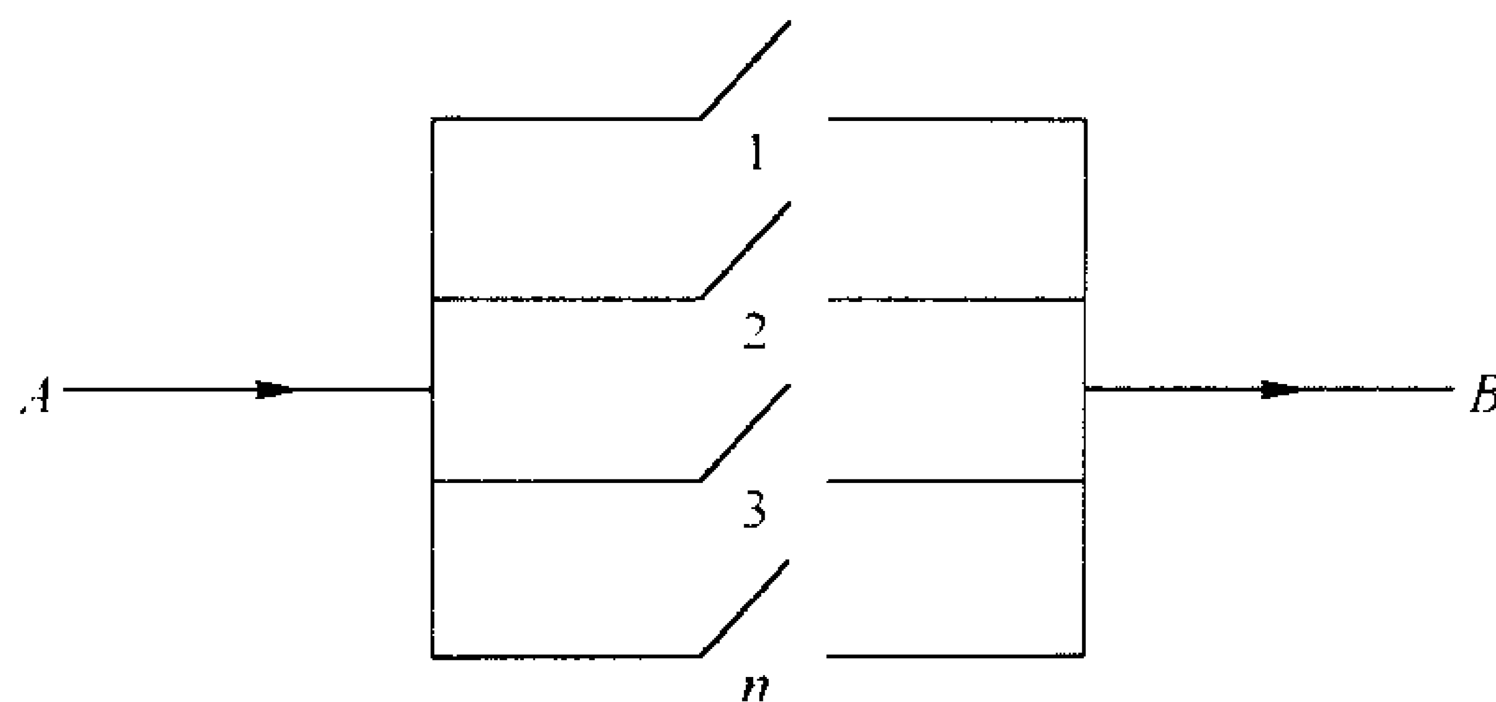


图 3-2 并联系统：假定电流从 A 流到 B

解 令 A_i 表示元件 i 工作的事件，则

$$\begin{aligned} P\{\text{系统工作}\} &= 1-P\{\text{系统不工作}\} \\ &= 1-P\{\text{每个元件都不工作}\} \\ &= 1-P\left(\bigcap_i A_i^c\right) \\ &= 1-\prod_{i=1}^{n}(1-p_i) \qquad \text{由独立性} \end{aligned}$$

■

87

例 4h 对于由掷两个均匀的骰子组成的独立重复试验而言，问两个骰子点数之和为 5 的结果出现在它们的点数之和为 7 的结果之前的概率是多少?

解 令 E_n 表示前 $n-1$ 次试验 5 点与 7 点都没有出现而第 n 次试验出现 5 点这一事件，那么所求的概率为

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

现在，因为在每一次试验中 5 点出现的概率为$\frac{4}{36}$，而 7 点出现的概率为$\frac{6}{36}$，所以由试验的独立性可得

$$P(E_n) = \left(1-\frac{10}{36}\right)^{n-1}\frac{4}{36}$$

从而有

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \frac{1}{9}\sum_{n=1}^{\infty}\left(\frac{13}{18}\right)^{n-1} = \frac{1}{9}\frac{1}{1-\frac{13}{18}} = \frac{2}{5}$$

此结果也可由条件概率求得，设 E 表示 5 点出现在 7 点之前这一事件，那么若把第一次试验的结果作为条件，则可得到所求的概率 $P(E)$. 解法如下：设 F 表示第一次试验结果为 5 点的事件，G 表示第一次试验结果为 7 点的事件，而 H 表示第一次试验结果既不是 5 点也不是 7 点的事件. 以这些事件之一出现作为条件给出

$$P(E) = P(E|F)P(F) + P(E|G)P(G) + P(E|H)P(H)$$

但是

$$P(E|F) = 1$$
$$P(E|G) = 0$$

$$P(E|H)=P(E)$$

上述前两个等式是显而易见的. 第三个等式是因为，若第一次试验结果既不是5点也不是7点，那么此时的情况正好与问题刚提出时一样；也就是说，试验者将继续掷这两个均匀的骰子，直到5点或7点出现为止. 此外，试验是相互独立的，因此第一次试验的结果不影响随后的各次试验. 由于 $P(F)=\frac{4}{36}$，$P(G)=\frac{6}{36}$，$P(H)=\frac{26}{36}$，我们得到

$$P(E)=\frac{1}{9}+P(E)\frac{13}{18}$$

从而

$$P(E)=\frac{2}{5}$$

读者会看到这个答案是相当直观的，这是因为在任何一次试验中，5点出现的概率为$\frac{4}{36}$，而7点出现的概率为$\frac{6}{36}$. 因此，可以直观地看出5点出现在7点之前的可能是4∶6，故所求的概率应为$\frac{4}{10}$，事实上的确如此.

88

用相同的推理可以证明，若 E 与 F 是某一试验的两个互不相容的事件，那么当独立地重复进行这一试验时，事件 E 发生在事件 F 之前的概率为

$$\frac{P(E)}{P(E)+P(F)}$$ ■

下一个例子在概率论的历史上占据着重要的地位，即著名的**分赌注问题**(problem of the points). 这个问题描述如下：两个赌徒下了注，然后玩某个赌博游戏，规定得胜者将获得所有赌注. 但在谁也没有得胜之前赌博因故中止了，此时两人的比分不同，那么这些赌注该怎么分?

这个问题是1654年德梅尔(de Méré)爵士(他当时是职业赌徒)向法国数学家帕斯卡(Pascal)提出的. 在攻克这一难题的过程中，帕斯卡提出了这样一个重要思想：赌徒应得的赌注的比例，取决于从现在开始比赛继续下去他们各自能取胜的概率. 帕斯卡解决了一些特殊情况. 更为重要的是，他与法国著名的数学家费马(Fermat)建立了联系，费马在数学界享有很高的声望. 他们之间通信的结果是，不仅完全解决了分赌注问题，而且还为解决有关机会游戏的许多其他问题奠定了基础. 有些人把他们建立联系的这一天看作概率论的"生日"，这对于激发欧洲数学家对概率论的兴趣也起着重要的作用，因为当时最重要的数学家都认识帕斯卡和费马. 例如，在他们建立联系的短时间内，年轻的荷兰数学家惠更斯(Huygens)曾来到巴黎讨论这些问题及其解法，人们对这个新领域的兴趣和积极性迅速高涨起来.

例4i(分赌注问题) 进行某独立重复试验，每一次成功的概率为 p，失败的概率为 $1-p$. 问在 m 次失败之前取得 n 次成功的概率是多少? 如果A、B二人这样赌：试验成功一次A得1点，试验失败一次B得1点，规定A得到 n 点取胜，B得到 m 点取胜，那么所求的概率正是A获胜的概率.

解 我们将给出两个解法，前一个归功于帕斯卡，后者归功于费马.

设 $P_{n,m}$ 表示 n 次成功发生在 m 次失败之前的概率. 以第一次试验的结果作为条件，我们得到

89

$$P_{n,m}=pP_{n-1,m}+(1-p)P_{n,m-1} \qquad n\geqslant 1, m\geqslant 1$$

(为什么? 试述理由.)运用明显的边界条件 $P_{n,0}=0$，$P_{0,m}=1$，由上述方程可以解出 $P_{n,m}$. 这里

不写出其冗长的细节，下面转而考虑费马的解法.

费马指出，为使 n 次成功发生在 m 次失败之前，其充要条件是在前 $m+n-1$ 次试验中至少有 n 次成功.（即使在 $m+n-1$ 次试验完成之前，胜负已定，我们仍可想象必要的试验仍在进行.）这是对的，因为在前 $m+n-1$ 次试验中至少成功 n 次，则在前 $m+n-1$ 次试验中至多失败 $m-1$ 次，于是 n 次成功发生在 m 次失败之前. 另一方面，若在前 $m+n-1$ 次试验中成功次数少于 n，那么这些试验中至少失败 m 次，因此在 m 次失败之前未获得 n 次成功.

这样，如在例 4f 中所指出的，$m+n-1$ 次试验中恰有 k 次成功的概率为 $\binom{m+n-1}{k}p^k(1-p)^{m+n-1-k}$，我们得到所求的 m 次失败之前取得 n 次成功的概率为

$$P_{n,m}=\sum_{k=n}^{m+n-1}\binom{m+n-1}{k}p^k(1-p)^{m+n-1-k}$$ ■

下一个例子是著名的赌徒输光问题. ⊖

例 4j（赌徒输光问题） A、B 两个赌徒用连续抛硬币的方法赌钱. 每抛一次，若出现正面则 A 赢 B 一元钱，若出现反面则 A 付给 B 一元钱. 他们要一直赌到某个人输光为止. 如果假定连续抛硬币是相互独立的，而且每一次出现正面的概率均为 p，若开始时 A 有 i 元，B 有 $N-i$元，求 A 赢得所有钱的概率.

解 设 E 表示从 A 有 i 元. B 有 $N-i$ 元开始 A 赢得所有钱这一事件，它显然与 A 开始时的钱数有关，故可令 $P_i=P(E)$. 以第一次抛硬币的结果作为条件，我们将得到 $P(E)$的一个表达式：以 H 表示第一次抛硬币出现正面这一事件，那么

$$P_i=P(E)=P(E|H)P(H)+P(E|H^c)P(H^c)=pP(E|H)+(1-p)P(E|H^c)$$

如果已知 H 已发生，那么第一次抛硬币后 A 有 $i+1$ 元、B 有 $N-(i+1)$元. 由于假定连续抛硬币是独立的，且出现正面的概率总是 p，所以在 H 已发生的条件下 A 赢得所有钱的概率，正好等于从 A 有 $i+1$ 元、B 有 $N-(i+1)$元开始赌时 A 赢得所有钱的概率. 因此有

90

$$P(E|H)=P_{i+1}$$

同理，

$$P(E|H^c)=P_{i-1}$$

因此，令 $q=1-p$，我们得到

$$P_i=pP_{i+1}+qP_{i-1}\qquad i=1,2,\cdots,N-1 \tag{4-2}$$

现运用明显的边界条件 $P_0=0$ 与 $P_N=1$ 来解方程(4-2). 由于 $p+q=1$，此方程等价于

$$pP_i+qP_i=pP_{i+1}+qP_{i-1}$$

或

$$P_{i+1}-P_i=\frac{q}{p}(P_i-P_{i-1})\qquad i=1,2,\cdots,N-1 \tag{4-3}$$

于是，因为 $P_0=0$，由式(4-3)可得

$$P_2-P_1=\frac{q}{p}(P_1-P_0)=\frac{q}{p}P_1$$

⊖ 本节其余内容为可选内容.

$$
\begin{aligned}
P_3-P_2&=\frac{q}{p}(P_2-P_1)=\left(\frac{q}{p}\right)^2P_1\\
&\vdots\\
P_i-P_{i-1}&=\frac{q}{p}(P_{i-1}-P_{i-2})=\left(\frac{q}{p}\right)^{i-1}P_1\\
&\vdots\\
P_N-P_{N-1}&=\frac{q}{p}(P_{N-1}-P_{N-2})=\left(\frac{q}{p}\right)^{N-1}P_1
\end{aligned}
\tag{4-4}
$$

将式(4-4)的前 $i-1$ 项相加得到

$$P_i-P_1=P_1\left[\left(\frac{q}{p}\right)+\left(\frac{q}{p}\right)^2+\cdots+\left(\frac{q}{p}\right)^{i-1}\right]$$

或

$$
P_i=\begin{cases}
\dfrac{1-\left(\dfrac{q}{p}\right)^i}{1-\left(\dfrac{q}{p}\right)}P_1 & \text{若}\ \dfrac{q}{p}\neq 1\\[2ex]
iP_1 & \text{若}\ \dfrac{q}{p}=1
\end{cases}
$$

91

运用条件 $P_N=1$，我们有

$$
P_1=\begin{cases}
\dfrac{1-\left(\dfrac{q}{p}\right)}{1-\left(\dfrac{q}{p}\right)^N} & \text{若}\ p\neq\dfrac{1}{2}\\[2ex]
\dfrac{1}{N} & \text{若}\ p=\dfrac{1}{2}
\end{cases}
$$

从而有

$$
P_i=\begin{cases}
\dfrac{1-\left(\dfrac{q}{p}\right)^i}{1-\left(\dfrac{q}{p}\right)^N} & \text{若}\ p\neq\dfrac{1}{2}\\[2ex]
\dfrac{i}{N} & \text{若}\ p=\dfrac{1}{2}
\end{cases}
\tag{4-5}
$$

令 Q_i 表示 A 有 i 元、B 有 $N-i$ 元开始赌，B 赢得所有钱的概率，那么由对称性，只需交换 p 与 q、i 与 $N-i$ 的位置，我们就得到

$$
Q_i=\begin{cases}
\dfrac{1-\left(\dfrac{p}{q}\right)^{N-i}}{1-\left(\dfrac{p}{q}\right)^N} & \text{若}\ q\neq\dfrac{1}{2}\\[2ex]
\dfrac{N-i}{N} & \text{若}\ q=\dfrac{1}{2}
\end{cases}
$$

但 $q=\frac{1}{2}$ 与 $p=\frac{1}{2}$ 等价，故当 $q\neq\frac{1}{2}$ 时有

$$P_i+Q_i=\frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N}+\frac{1-\left(\frac{p}{q}\right)^{N-i}}{1-\left(\frac{p}{q}\right)^N}=\frac{p^N-p^N\left(\frac{q}{p}\right)^i}{p^N-q^N}+\frac{q^N-q^N\left(\frac{p}{q}\right)^{N-i}}{q^N-p^N}$$

$$=\frac{p^N-p^{N-i}q^i-q^N+q^ip^{N-i}}{p^N-q^N}=1$$

当 $p=q=\frac{1}{2}$ 时此结果仍成立，所以有

$$P_i+Q_i=1$$

总之，这个等式表明，A、B 中某一人将赢得所有钱的概率为 1；或者说，A 的钱总在 1 与 $N-1$ 之间而赌博无休止地进行下去的概率为 0.（读者必须注意，本来这场赌博有三个可能结果，而不是两个，即 A 胜、B 胜和谁也不胜地永远进行下去，我们刚才证明了最后一个事件的概率为 0.）

现在用具体数字对上述结果进行说明. 若开始赌时 A 有 5 元，B 有 10 元，则当 p 为 0.5 时，A 胜的概率为 $\frac{1}{3}$. 而当 p 为 0.6 时，A 胜的概率增为

92

$$\frac{1-\left(\frac{2}{3}\right)^5}{1-\left(\frac{2}{3}\right)^{15}}\approx 0.87$$

赌徒输光问题有一个特例，也称为**赌博持续时间**(duration of play)的问题. 这个问题是 1657 年法国数学家费马向荷兰数学家惠更斯提出的，后来被惠更斯解决. 该问题如下：设 A、B 各有 12 枚硬币，他们以掷三个骰子的方式赌这些钱. 若出现 11 点(谁掷这些骰子都可以)，则 A 给 B 一枚硬币；若出现 14 点，则 B 给 A 一枚硬币. 谁先赢得所有的硬币就获胜. 因为 $P\{\text{掷出 11 点}\}=\frac{27}{216}$，而 $P\{\text{掷出 14 点}\}=\frac{15}{216}$，故由例 4h 可知，对 A 而言，这正是 $p=\frac{15}{42}$，$i=12$，$N=24$ 时的赌徒输光问题(例 4j). 一般的赌徒输光问题由数学家詹姆士 · 伯努利(James Bernoulli)解出，而这个结果是在他去世 8 年后于 1713 年发表的.

赌徒输光问题的一个应用是药品检验问题. 假设为治某种病研制两种新药，药品 i 的治愈率为 $P_i(i=1,2)$，即患者服用药品 i 得到治愈的概率为 P_i. 然而，治愈率是未知的，我们在考虑一种方法决定 $P_1>P_2$ 或 $P_2>P_1$. 为做出选择，考虑如下试验：数对患者同时治疗，每一对中每人分别服用药品 1 和药品 2. 每种药品的结果是固定的，对于固定的、预先确定的数目，若一种药品治愈的累积数目超过另一种药品治愈的累积数目，则试验停止. 更正式地，令

$$X_j=\begin{cases}1 & \text{服用药品 1 的第 } j \text{ 对患者治愈}\\ 0 & \text{其他}\end{cases}$$

$$Y_j=\begin{cases}1 & \text{服用药品 2 的第 } j \text{ 对患者治愈}\\ 0 & \text{其他}\end{cases}$$

对预先给定的正整数 M，试验 N 对后停止，其中 N 是 n 的第一个值，使得

$$X_1+\cdots+X_n-(Y_1+\cdots+Y_n)=M$$

或

$$X_1+\cdots+X_n-(Y_1+\cdots+Y_n)=-M$$

经验证前者 $P_1>P_2$，后者 $P_2>P_1$.

为确定上一试验是否是好试验，我们想知道导致一个错误决定的概率. 即对给定的 P_1 和 P_2，其中 $P_1>P_2$，试验错误地断言 $P_2>P_1$ 的概率是多少? 为求此概率，每对试验检验完之后，服用药品 1 和药品 2 的累积治愈差别以概率 $P_1(1-P_2)$上升到 1——因为药品 1 能够治愈，而药品 2 不能，或者以概率$(1-P_1)P_2$ 下降，或者以概率 $P_1P_2+(1-P_1)(1-P_2)$保持不变. 因此，我们考虑累积差别的不同变化，则差别以概率

93

$$P=P\{\text{上升到 }1\mid\text{上升到 }1\text{ 或下降到 }1\}=\frac{P_1(1-P_2)}{P_1(1-P_2)+(1-P_1)P_2}$$

上升到 1，而以概率

$$1-P=\frac{P_2(1-P_1)}{P_1(1-P_2)+(1-P_1)P_2}$$

下降到 1.

因此试验断言 $P_2>P_1$ 的概率等价于以概率 P 赢得一元钱的赌徒在上升 M 之前下降 M 的概率. 由式(4-5)，当 $i=M,N=2M$ 时，概率如下：

$$P\{\text{试验断言 }P_2>P_1\}=1-\frac{1-\left(\frac{1-P}{P}\right)^M}{1-\left(\frac{1-P}{P}\right)^{2M}}=1-\frac{1}{1+\left(\frac{1-P}{P}\right)^M}=\frac{1}{1+\gamma^M}$$

其中，

$$\gamma=\frac{P}{1-P}=\frac{P_1(1-P_2)}{P_2(1-P_1)}$$

例如，若 $P_1=0.6$，$P_2=0.4$，当 $M=5$ 时，做出错误判断的概率是 0.017，而 $M=10$ 时概率降为 0.000 3. ■

假定考虑一个元素的集合并确定这个集合中至少有一个元素具有某种性质. 我们可以通过在这个集合中随机地选择一个元素，使得被选择的每个元素具有正的概率，把该问题化为概率问题. 那么，原来的问题可以通过考虑随机选择的元素具有某种性质的概率而得到回答. 假若概率是正的，那么我们已经确定了集合中至少有一个元素具有某种性质，若概率为零，那么集合中没有一个元素具有某种性质.

94

例 4k 由 n 个顶点组成的完全图定义为平面中的 n 点(称为顶点)集，连接每对顶点组成 $\binom{n}{2}$条线(称为边). 图 3-3 给出有 3 个顶点的完全图. 假定由 n 个顶点组成的完全图的每条边染成红色或蓝色. 对固定的整数 k，考虑能否将 k 个顶点的所有$\binom{k}{2}$条边染成同色. 这是一个概率问题，当 n 不是太大时，答案是肯定的.

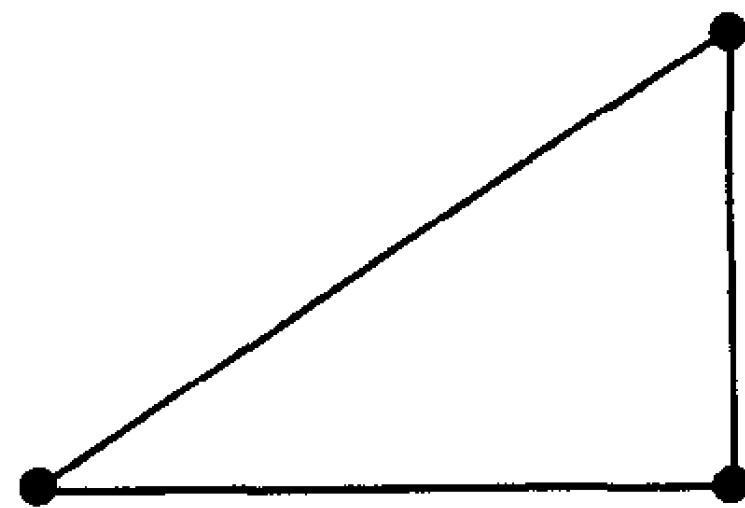


图 3-3

假设每条边染成红色或蓝色是独立的，即每条边染成红色的概率是$\frac{1}{2}$. 对 k 个顶点的所有$\binom{n}{k}$个集，定义事件 $E_i\left(i=1,2,\cdots,\binom{n}{k}\right)$如下：

$$E_i=\{k\text{ 个顶点的第 }i\text{ 个集的所有边染成同色}\}$$

由于 k 个顶点的所有 $\binom{k}{2}$ 条边等可能地染成红色或蓝色，所以染成同色的概率为

$$P(E_i)=2\left(\frac{1}{2}\right)^{k(k-1)/2}$$

因为

$$P\left(\bigcup_i E_i\right)\leqslant \sum_i P(E_i) \qquad \text{（布尔不等式）}$$

我们得到 $P\left(\bigcup_i E_i\right)$，$k$ 个顶点的所有连接边被染成同色的概率满足

$$P\left(\bigcup_i E_i\right)\leqslant \binom{n}{k}\left(\frac{1}{2}\right)^{\frac{k(k-1)}{2}-1}$$

95

因此，若

$$\binom{n}{k}\left(\frac{1}{2}\right)^{\frac{k(k-1)}{2}-1}<1$$

或等价地，若

$$\binom{n}{k}<2^{\frac{k(k-1)}{2}-1}$$

则 k 个顶点的 $\binom{n}{k}$ 个集中至少一个集的所有连接边染成同色的概率小于 1. 因此，在上述的 n 和 k 的条件下，由此推出 k 个顶点的集合中没有一个集的所有连接边染成同色是一个正概率. 换句话说，至少有一种染边的方法使得 k 个顶点的集合中其所有连接边的颜色不完全相同. ■

注 (a) 值得注意的是，上述建立在 n 和 k 上的结论保证了满足所希望的性质的染色方法的存在性，但是对如何得到这种方法却没有给出任何信息.（一种可能的方法是随机地染色，然后验证染色的结果是否满足要求，重复这个过程直到满足为止.）

(b) 将概率引入到纯确定性的问题中的方法称为概率方法(probabilistic method). [㊀] 这个方法的其他例子将在第 7 章的理论练习 22 和例 2q 中给出.

3.5 $P(\cdot\,|\,F)$是一种概率

条件概率具备普通概率的所有基本性质，这由命题 5.1 所证明，此命题说明了条件概率 $P(E|F)$满足概率的三条公理.

命题 5.1

(a) $0\leqslant P(E|F)\leqslant 1$.

(b) $P(S|F)=1$.

(c) 若 $E_i(i=1,2,\cdots)$是互不相容的事件，则

96

$$P\left(\bigcup_1^{\infty} E_i\,|\,F\right)=\sum_1^{\infty}P(E_i\,|\,F)$$

㊀ 参见 N. Alon, J. Spencer, and P. Erdos, *The Probabilistic Method* (New York: John Wiley & Sons, Inc., 1992).

证明 为证(a)，必须证明 $0\leqslant P(EF)/P(F)\leqslant 1$，不等式的左边显然成立，不等式的右边成立是因为 $EF\subset F$ 蕴涵着 $P(EF)\leqslant P(F)$.

(b)成立是因为

$$P(S|F)=\frac{P(SF)}{P(F)}=\frac{P(F)}{P(F)}=1$$

(c)可由如下推导得出：

$$\begin{aligned}P\left(\bigcup_{i=1}^{\infty} E_i \,\middle|\, F\right) &= \frac{P\left(\left(\bigcup_{i=1}^{\infty} E_i\right)F\right)}{P(F)} \\ &= \frac{P\left(\bigcup_1^{\infty} E_iF\right)}{P(F)} \qquad \text{因为}\left(\bigcup_1^{\infty} E_i\right)F=\bigcup_1^{\infty} E_iF \\ &= \frac{\sum_1^{\infty} P(E_iF)}{P(F)} = \sum_1^{\infty} P(E_i|F)\end{aligned}$$

倒数第二个等式成立是因为 $E_iE_j=\varnothing$ 蕴涵着 $E_iFE_jF=\varnothing$. ■

如果定义 $Q(E)=P(E|F)$，则由命题 5.1 知，$Q(E)$ 可看作定义于 S 中的事件上的概率函数. 于是，先前对概率所证明的一切命题都适用于 $Q(E)$. 例如，我们有

$$Q(E_1\cup E_2)=Q(E_1)+Q(E_2)-Q(E_1E_2)$$

或等价地，

$$P(E_1\cup E_2|F)=P(E_1|F)+P(E_2|F)-P(E_1E_2|F)$$

若再定义条件概率 $Q(E_1|E_2)=Q(E_1E_2)/Q(E_2)$，则由式(3-1)可得到

$$Q(E_1)=Q(E_1|E_2)Q(E_2)+Q(E_1|E_2^c)Q(E_2^c) \tag{5-1}$$

因为

97

$$Q(E_1|E_2)=\frac{Q(E_1E_2)}{Q(E_2)}=\frac{P(E_1E_2|F)}{P(E_2|F)}=\frac{P(E_1E_2F)/P(F)}{P(E_2F)/P(F)}=P(E_1|E_2F)$$

所以式(5-1)等价于

$$P(E_1|F)=P(E_1|E_2F)P(E_2|F)+P(E_1|E_2^cF)P(E_2^c|F)$$

例 5a 回顾例 3a，某保险公司认为，人可分为两类——易出事故的和比较谨慎的. 在任何给定的一年期间，易出事故的人将发生一起事故的概率为 0.4，而对比较谨慎的人而言这个概率是 0.2. 若已知某新保险客户在第一年已经出过一次事故，问他在保险有效的第二年又出一次事故的条件概率是多少？

解 令 A 表示这个保险客户是易出事故的人这一事件，而 $A_i(i=1,2)$ 表示他在第 i 年出一次事故，那么以他是不是易出事故的人为条件，可以算出所求的概率 $P(A_2|A_1)$ 如下：

$$P(A_2|A_1)=P(A_2|AA_1)P(A|A_1)+P(A_2|A^cA_1)P(A^c|A_1)$$

而

$$P(A|A_1)=\frac{P(A_1A)}{P(A_1)}=\frac{P(A_1|A)P(A)}{P(A_1)}$$

但是，在例 3a 中已知 $P(A)=0.3$，且得到 $P(A_1)=0.26$，于是有

$$P(A|A_1)=\frac{(0.4)(0.3)}{0.26}=\frac{6}{13}$$

从而

$$P(A^c|A_1)=1-P(A|A_1)=\frac{7}{13}$$

由于 $P(A_2|AA_1)=0.4$，$P(A_2|A^cA_1)=0.2$，因此

$$P(A_2|A_1)=(0.4)\frac{6}{13}+(0.2)\frac{7}{13}\approx 0.29$$ ■

下一个例子讨论游程论中的一个问题.

例 5b 进行一个独立重复试验，每一次试验成功的概率为 p，失败的概率为 $q=1-p$. 我们的兴趣在于计算一个由 n 次连续成功试验组成的游程出现在由 m 次连续失败试验组成的游程之前的概率.

解 以 E 表示 n 次连续成功试验组成的游程出现在 m 次连续失败试验组成的游程之前这一事件，为求 $P(E)$，先以第一次试验的结果作为条件，设 H 表示第一次试验结果为成功的事件，我们得到

$$P(E)=pP(E|H)+qP(E|H^c) \tag{5-2}$$

98

现在，已知第一次试验是成功的，那么，若此后 $n-1$ 次试验的结果都是成功的，则这就是我们能得到 m 次连续失败试验组成的游程之前有 n 次连续成功试验组成的游程的一种方式. 因此，我们又可以把这件事是否发生作为条件，就是说，令 F 表示从第 2 次到第 n 次试验都成功这一事件，可得到

$$P(E|H)=P(E|FH)P(F|H)+P(E|F^cH)P(F^c|H) \tag{5-3}$$

显然，$P(E|FH)=1$；另一方面，若事件 F^cH 发生，则第一次试验是成功的，此后 $n-1$ 次试验必有某一次失败. 但当这一失败发生时，在它之前的成功全部作废了，因而情况正相当于从失败开始，故得

$$P(E|F^cH)=P(E|H^c)$$

由于重复试验的独立性蕴涵着 F 与 H 独立，还由于 $P(F)=p^{n-1}$，据式(5-3)可得到

$$P(E|H)=p^{n-1}+(1-p^{n-1})P(E|H^c) \tag{5-4}$$

同理可得 $P(E|H^c)$的一个表达式. 也就是说，设 G 表示从第 2 次到第 m 次试验全失败的事件，那么有

$$P(E|H^c)=P(E|GH^c)P(G|H^c)+P(E|G^cH^c)P(G^c|H^c) \tag{5-5}$$

GH^c 表示前 m 次试验全失败的事件，故 $P(E|GH^c)=0$. 还有，若 G^cH^c 发生，则第一次试验失败了，而且在此后的 $m-1$ 次试验中至少有一次成功. 这样，由于这一次成功，在它前面的所有失败全部作废了，故有

$$P(E|G^cH^c)=P(E|H)$$

于是，因为 $P(G^c|H^c)=P(G^c)=1-q^{m-1}$，故式(5-5)可化为

$$P(E|H^c)=(1-q^{m-1})P(E|H) \tag{5-6}$$

解方程(5-4)与(5-6)得

$$P(E|H)=\frac{p^{n-1}}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}}$$

和

$$P(E|H^c)=\frac{(1-q^{m-1})p^{n-1}}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}}$$

因此

$$P(E)=pP(E|H)+qP(E|H^c)=\frac{p^n+qp^{n-1}(1-q^{m-1})}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}}$$

99

$$=\frac{p^{n-1}(1-q^m)}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}} \tag{5-7}$$

由问题的对称性知，将式(5-7)中的 m 与 n、p 与 q 交换位置，就可以得到 m 次连续失败试验组成的游程出现在 n 次连续成功试验组成的游程之前的概率，于是有

$P\{m$ 次连续失败试验组成的游程出现在 n 次连续成功试验组成的游程之前$\}$

$$=\frac{q^{m-1}(1-p^n)}{q^{m-1}+p^{n-1}-q^{m-1}p^{n-1}} \tag{5-8}$$

因为式(5-7)与式(5-8)右边之和为 1，因此得到最终将出现要么 n 次连续成功试验组成的游程要么 m 次连续失败试验组成的游程的概率为 1.

作为式(5-7)的一个例子，注意当抛一枚均匀硬币时，2 个正面的游程出现在 3 个反面的游程之前的概率为$\frac{7}{10}$，而 2 个连续正面先于 4 个连续反面的概率则增长到$\frac{5}{6}$. ■

下一个例子再次研究配对问题(第 2 章例 5m)，这里运用条件概率求解.

例 5c 在一次聚会上，n 个人摘掉他们的帽子，然后把这些帽子混放到一起，每人随机地选择一顶，若某人选中了他自己的帽子，我们就说出现了一个配对.

(a) 没有配对的概率是多少?

(b) 恰有 k 个配对的概率是多少?

解 (a) 以 E 表示没有配对这一事件，它显然与 n 有关，故可记为 $P_n=P(E)$. 以第一个人选中自己的帽子或没有选中自己的帽子为条件，分别记这两个事件为 M 与 M^c，从而

$$P_n=P(E)=P(E|M)P(M)+P(E|M^c)P(M^c)$$

显然，$P(E|M)=0$，因此

$$P_n=P(E|M^c)\frac{n-1}{n} \tag{5-9}$$

$P(E|M^c)$是在已知 $n-1$ 个人中的某个人(此人的帽子已被第一个人选走)必定选不到自己的帽子的条件下，这 $n-1$ 个人选 $n-1$ 顶帽子没有配对的概率. 这有两种互不相容的选取方式：要么额外的人(即被第一个人选走帽子的人)没选上额外的(即第一个人的)帽子且其余的人中也没有配对，要么额外的人选中了额外的帽子且其余的人中也没有配对. 前者的概率是 P_{n-1}，后者的概率是$[1/(n-1)]P_{n-2}$，我们有

$$P(E|M^c)=P_{n-1}+\frac{1}{n-1}P_{n-2}$$

100

于是式(5-9)化为

$$P_n=\frac{n-1}{n}P_{n-1}+\frac{1}{n}P_{n-2}$$

或等价地，有

$$P_n-P_{n-1}=-\frac{1}{n}(P_{n-1}-P_{n-2}) \tag{5-10}$$

但是，由于 P_n 是 n 个人在他们的帽子里任选一顶没有配对的概率，我们有

$$P_1=0 \qquad P_2=\frac{1}{2}$$

由式(5-10)可得

$$P_3-P_2=-\frac{1}{3}(P_2-P_1)=-\frac{1}{3!} \quad 或 \quad P_3=\frac{1}{2!}-\frac{1}{3!}$$

$$P_4-P_3=-\frac{1}{4}(P_3-P_2)=\frac{1}{4!} \quad 或 \quad P_4=\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}$$

一般地，有

$$P_n=\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-\cdots+\frac{(-1)^n}{n!}$$

(b) 为计算恰好有 k 个配对的概率，先考虑任意固定的 k 个人. 他们且只有他们选中自己帽子的概率为

$$\frac{1}{n}\frac{1}{n-1}\cdots\frac{1}{n-(k-1)}P_{n-k}=\frac{(n-k)!}{n!}P_{n-k}$$

其中，P_{n-k}是已知 k 个人选中自己的帽子后，其余 $n-k$ 个人在他们自己的帽子里选取而没有配对的条件概率. 因为这 k 个人有$\binom{n}{k}$种选法，故所求的恰好有 k 个配对的概率为

$$\frac{P_{n-k}}{k!}=\frac{\frac{1}{2!}-\frac{1}{3!}+\cdots+\frac{(-1)^{n-k}}{(n-k)!}}{k!}$$

■

概率论的另一个重要概念是事件的条件独立性. 我们称事件 E_1 与 E_2 对于给定的 F 是*条件独立的*，如果在已知 F 已发生的条件下，E_1 发生的概率不因 E_2 是否发生而改变. 确切地说，E_1 与 E_2 称为关于 F 是条件独立的，如果

$$P(E_1|E_2F)=P(E_1|F) \tag{5-11}$$

或等价地，

$$P(E_1E_2|F)=P(E_1|F)P(E_2|F) \tag{5-12}$$

101

条件独立的概念很容易推广到多于两个事件的情形，读者可以自己验证.

读者会发现，条件独立性的概念在例 5a 中已用过了，在那里，我们假定已知保险客户是或者不是易出事故的人之后，他在第 $i(i=1,2,\cdots)$ 年出一次事故的事件是条件独立的. [在取 $P(A_2|AA_1)=0.4$，$P(A_2|A^cA_1)=0.2$ 时用到这一点.]下一个例子进一步解释条件独立性的概念，有时称为拉普拉斯继承准则.

例 5d(拉普拉斯继承准则) 盒中有 $k+1$ 枚硬币. 抛第 i 枚硬币时，其出现正面的概率是 i/k，$i=0, 1, 2, \cdots, k$. 从盒中随机取出一枚硬币，并反复地抛，若前 n 次的结果皆为正面，问第 $n+1$ 次仍抛出正面的概率是多少?

解 令 C_i 为开始取出的是第 $i(i=0,1,2,\cdots,k)$ 枚硬币这一事件；F_n 为前 n 次都抛出正面的事件，H 为第 $n+1$ 次抛出正面的事件，所求的概率 $P(H|F_n)$可表示为

$$P(H|F_n)=\sum_{i=0}^{k}P(H|F_nC_i)P(C_i|F_n)$$

若取出的是第 i 枚硬币，则有理由假定各次抛出的结果条件独立，且出现正面的概率是 i/k. 于是有

$$P(H|F_nC_i)=P(H|C_i)=\frac{i}{k}$$

再者，

$$P(C_i|F_n)=\frac{P(C_iF_n)}{P(F_n)}=\frac{P(F_n|C_i)P(C_i)}{\sum_{j=0}^{k}P(F_n|C_j)P(C_j)}=\frac{(i/k)^n[1/(k+1)]}{\sum_{j=0}^{k}(j/k)^n[1/(k+1)]}$$

因此有

$$P(H|F_n)=\frac{\sum_{i=0}^{k}(i/k)^{n+1}}{\sum_{j=0}^{k}(j/k)^n}$$

102

但当 k 足够大时，可利用积分近似

$$\frac{1}{k}\sum_{i=0}^{k}\left(\frac{i}{k}\right)^{n+1}\approx\int_0^1 x^{n+1}\mathrm{d}x=\frac{1}{n+2}$$

$$\frac{1}{k}\sum_{j=0}^{k}\left(\frac{j}{k}\right)^{n}\approx\int_0^1 x^{n}\mathrm{d}x=\frac{1}{n+1}$$

从而对很大的 k 有

$$P(H|F_n)\approx\frac{n+1}{n+2}$$ ■

小结

对于事件 E 与 F，若给定 F，E 发生的条件概率 $P(E|F)$ 定义为

$$P(E|F)=\frac{P(EF)}{P(F)}$$

乘法公式为

$$P(E_1E_2\cdots E_n)=P(E_1)P(E_2|E_1)\cdots P(E_n|E_1E_2\cdots E_{n-1})$$

可以利用 F 是否发生作为条件计算 $P(E)$，得到如下恒等式：

$$P(E)=P(E|F)P(F)+P(E|F^c)P(F^c)$$

$P(H)|P(H^c)$ 称为事件 H 的优势率. 恒等式

$$\frac{P(H|E)}{P(H^c|E)}=\frac{P(H)P(E|H)}{P(H^c)P(E|H^c)}$$

表明，如果得到新证据 E，则 H 的新优势率等于其旧优势率乘以新证据使 H 成立的条件下的条件概率与新证据 H 不成立的条件下的条件概率之比.

令 $F_i(i=1,2,\cdots,n)$ 是互不相容的事件，$\bigcup_{i=1}^{n}F_i=S$(即整个样本空间). 等式

$$P(F_j|E)=\frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n}P(E|F_i)P(F_i)}$$

称为贝叶斯公式. 若事件 $F_i(i=1,\cdots,n)$ 为竞争假设，则贝叶斯公式说明如何计算在给定 E 的条件下的条件概率.

103

若 $P(EF)=P(E)P(F)$，则称事件 E 与 F 是独立的，该条件等价于 $P(E|F)=P(E)$ 与

$P(F|E)=P(F)$. 这样，若事件 E 或 F 的发生不影响彼此的概率，则称事件 E 与 F 是独立的.

事件 $E_1,E_2,\cdots,E_n$ 称为是独立的，若对其任意子集 $E_{i_1},\cdots,E_{i_r}$ 有

$$P(E_{i_1}\cdots E_{i_r})=P(E_{i_1})\cdots P(E_{i_r})$$

对固定的事件 F，可把 $P(E|F)$看成是定义于样本空间的事件 E 上的概率函数.

习题

1. 掷两个均匀的骰子，如果已知它们的点数不相同，问至少有一个是 6 点的条件概率是多少?
2. 掷两个均匀的骰子，已知点数之和为 $i(i=2\sim12)$，问第一个骰子是 6 点的条件概率是多少?
3. 利用式(2-1)计算，在一副桥牌中，如果南北两家共有 8 张黑桃，问东家有 3 张黑桃的条件概率是多少?
4. 已知两个均匀的骰子点数之和为 $i(i=2\sim12)$，问至少有一个骰子是 6 点的概率是多少?
5. 箱中有 6 个白球与 9 个黑球，不放回地从箱中随机取出 4 个球，问前两个是白球后两个是黑球的概率是多少?
6. 箱中有 12 个球，其中白球 8 个，现不放回地取出 4 个球. 已知取出的球中恰有 3 个白球，问取出的第一个与第三个球是白球的概率是多少? 如果每取出一球后都放回呢?
7. 国王出身于一个有两个孩子的家庭，另一个孩子是国王的姐妹的概率是多少?
8. 某夫妇有两个孩子，若年长的是女孩，那么两个孩子都是女孩的概率是多少?
9. 考虑 3 个箱子，A 箱中有 2 个白球与 4 个红球，B 箱中有 8 个白球与 4 个红球，C 箱中有 1 个白球与 3 个红球. 如果从每箱中各取出一个球，已知这三个球中正好有 2 个白球，问从 A 箱中取出白球的概率是多少?
10. 从一副 52 张扑克牌中随机地、不放回地取出 3 张，已知第二张与第三张都是黑桃，试求第一张也是黑桃的条件概率.
11. 吸烟的孕妇出现胎位不正是不吸烟孕妇的两倍，若 32%的分娩妇女是吸烟者，问有百分之多少的出现胎位不正的妇女是吸烟者?
12. 98%的婴儿分娩存活，而 15%的婴儿是剖腹产，其中剖腹产婴儿的存活率为 96%，问若某孕妇顺产，则她的婴儿存活的概率是多少?
13. 某社区 36%的家庭养狗，22%的家庭既养狗又养猫，而 30%的家庭养猫.
    (a) 随机选出一个家庭，求既养狗又养猫的概率.
    (b) 随机选出一个家庭，已知养猫，求又养狗的条件概率.

[104]

14. 某城市 46%的选民是独立党人，30%的选民是自由党人，24%的选民是保守党人. 在当地最新一次选举中，有 35%的独立党人、62%的自由党人、58%的保守党人参加投票. 随机选出一个选民，已知他参与了这次选举.
    (a) 求他是独立党人的概率.
    (b) 求他是自由党人的概率.
    (c) 求他是保守党人的概率.
    (d) 有百分之多少的选民参加了这次选举?
15. 参加戒烟班的 48%的女性和 37%的男性在一年后仍不吸烟，年底他们将参加庆功会. 已知该班最初有 62%是男性.
    (a) 问参加庆功会的百分之多少是女性?
    (b) 问最初班里有百分之多少的人参加了年底的庆功会?
16. 某大学有 52%的女生，5%的学生专业为计算机科学，其中计算机科学专业的 2%是女生. 如果随机选出一名学生，试求下列概率.
    (a) 已知该学生是计算机科学专业，问她是一个女生的概率.

(b) 已知该学生是一个女生，问她是计算机科学专业的学生的概率.

17. 调查 500 对双职工夫妇的年薪，有如下信息：

妻　子	丈　夫	
	少于 25 000 美元	多于 25 000 美元
少于 25 000 美元	212	198
多于 25 000 美元	36	54

例如，有 36%的夫妇，其中妻子的年薪多于 25 000 美元而丈夫的年薪少于 25 000 美元. 随机选出一对夫妇，试求下列概率：

(a) 丈夫的年薪少于 25 000 美元的概率.

(b) 如果丈夫的年薪多于 25 000 美元，妻子的年薪也多于这个数目的条件概率.

(c) 如果丈夫的年薪多于 25 000 美元，妻子的年薪少于这个数目的条件概率.

18. 一名大学毕业生打算参加夏天的保险考试的前三门，6 月她参加第一门考试，若通过这门考试，则她会参加 7 月的第二门考试，若再通过这门考试，则她会参加 9 月的第三门考试. 如果没有通过考试，则不允许她参加接下来的考试. 她通过第一门考试的概率为 0.9. 若已知她通过第一门考试，则通过第二门考试的条件概率是 0.8. 若已知她通过前两门考试，则通过第三门考试的条件概率是 0.7.

(a) 问她通过所有三门考试的概率是多少?

105

(b) 若她没有通过所有三门考试，问她没有通过第二门考试的条件概率是多少?

19. 一副 52 张牌(包含 4 个 A)随机地平均分给 4 个人. 每人都有 A 的概率为 p. 令 E_i 为第 i 个玩家恰好有一个 A 的事件，试利用乘法公式计算 $p=P(E_1E_2E_3E_4)$.

20. 箱中原有 5 个白球与 7 个黑球. 每次从箱中取出一个球，看看什么颜色，然后再添上 2 个与它同色的球一起放回箱中.

(a) 计算取出的前两个是黑球、随后的两个是白球的概率.

(b) 计算在取出的前 4 个球中恰有 2 个黑球的概率.

21. Ⅰ箱中有 2 个白球与 4 个红球，Ⅱ箱中有 1 个白球与一个红球. 随机地从Ⅰ箱中取出一个球放入Ⅱ箱，然后从Ⅱ箱中随机取出一个球.

(a) 求从Ⅱ箱中取出的球是白球的概率.

(b) 若从Ⅱ箱中取出的球是白球，求从Ⅰ箱转移到Ⅱ箱的那个球也是白球的条件概率.

22. 有 2 个球，每个球或者涂上黑色或者涂上金黄色，然后放入箱中. 假设每个球涂上黑色的概率均为 $\frac{1}{2}$，且这些事件是独立的.

(a) 假设已知金黄色已用过了(即至少有一个球涂上金黄色)，求两个球都涂上金黄色的条件概率.

(b) 现设箱子翻倒了，出来一个金黄色球，此时两个球都是金黄色的概率是多少? 解释所得的结果.

23. 要估计居住在某个有 100 000 人口城市中的年龄超过 50 岁的人数，有人提出采用如下方法："当你在街上散步时，数一数你遇到的超过 50 岁的人数，再算出他们占你遇到的人的百分数. 这样做几天以后，用 100 000 去乘得到的百分数就是所求的估值". 试评价这个方法.

提示：设这个城市中超过 50 岁的人所占比例为 p. 另外，令 α_1 表示一个 50 岁以下的人消耗在街上的时间所占的比例，α_2 表示一个超过 50 岁的人消耗在街上的时间所占的比例. 这个方法估计的是什么? 什么时候这个估值近似等于 p?

24. 假设 5%的男人与 0.25%的女人是色盲，且男人总数与女人总数相等. 现随机选出一人，发现是个色盲，此人是男人的概率是多少? 如果在居民中男人总数是女人总数的两倍，这个概率又是多少?

25. 某公司开车上班的职工把车停在停车场分给本公司的地方. 公司想估计开车上班的平均人数. 下列哪种方法能够用来估计? 给出解释.

(a) 随机选出 n 个职工，求出有多少人开车上班，然后对这 n 个值取平均.

(b) 随机在停车场选出 n 辆车，求出有多少人开这些车，然后对这 n 个值取平均.

26. 从一副洗好的 52 张牌中，每次取出一张牌直到第一个 A 出现. 若第 20 张牌出现第一个 A，试求下列条件概率，若下一张牌是

(a) 黑桃 A；(b)梅花 2.

27. 盒内有 15 个乒乓球，其中 9 个从未使用过. 随机取出 3 个球，用过后又放回盒里. 随后又从盒里取出 3 个球，试求这些球从未用过的概率.

106

28. 有两个盒子，第一盒中有黑、白弹子各一粒，第二个盒中有 2 粒黑的、1 粒白的弹子. 现随机选择一盒，再从所选的盒中随机取出一粒弹子，此弹子是黑色的概率是多少？若已知取出的是白弹子，它是从第一盒中取出的概率是多少？

29. 对“严格”这个词，英国人与美国人的拼写方法分别为“rigour”与“rigor”. 暂住在巴黎某宾馆中的一个人写了这个词，从他写的这个词随机地取出一个字母，发现是元音. 如果住在此宾馆里讲英语的人中，有 40%是英国人，其余 60%是美国人，问写字者是英国人的概率是多少？

30. 在例 3f 中，假定新证据满足不同可能的解释，事实上，只有 90%的把握说明罪犯拥有那些特征. 这种情形下，嫌疑犯有罪的概率是多少？(假定嫌疑犯有这些特征.)

31. 某班有 30 名学生，其中 15 名优秀生，10 名中等生，5 名差生. 另一班也有 30 名学生,其中 5 名优秀生，10 名中等生，15 名差生. 你(专家)知道这一信息，但分不清楚这两个班. 如果你随机从两个班各抽查一名学生，发现来自 A 班的学生是中等生，B 班的学生是差生. 试问 A 班比 B 班好的概率是多少？

32. A、B、C 三个商店分别有 50、75、100 名雇员，其中女雇员依次占 50%、60%、70%. 在所有的雇员中辞职是等可能的，与性别也无关. 现知一个雇员辞职了，还知道她是女雇员，问此人原在 C 商店工作的概率是多少？

33. (a) 一个赌徒的衣袋中放着两枚硬币，其中一枚是均匀的，另一枚两面都是正面. 他随机取一枚抛出，结果出现正面，问取出的是均匀硬币的概率是多少？

(b) 如果把这枚硬币再抛一次，结果又出现正面，问取出的是均匀硬币的概率是多少？

(c) 如果把这枚硬币再抛第三次，结果出现反面，问取出的是均匀硬币的概率是多少？

34. A 箱中有 5 个白球与 7 个黑球，B 箱中有 3 个白球与 12 个黑球. 抛一枚均匀的硬币，若抛出正面则从 A 箱中取出一球，若抛出反面则从 B 箱中取出一球. 现知取出的是一个白球，问抛这枚硬币时出现反面的概率是多少？

35. 在例 3a 中，已知保险客户在第一年没出事故，他在第二年出事故的概率是多少？

36. 由 3 个球组成的样本按如下方式取出：起初箱中有 5 个白球与 7 个红球，每次从箱中取出一球，记下它的颜色，然后再加入一个同样颜色的球一起放回箱中，试求此样本正好包含(a) 0 个白球；(b) 1 个白球；(c) 3 个白球；(d) 2 个白球的概率.

37. 将一副纸牌洗好后分为两份，每份 26 张. 从第一份中任取一张牌，发现是个“A”，然后把它放入第二份中，第二份牌洗好后再从中取出一张，求这张牌又是“A”的概率.

提示：以第二次取出的牌是不是从第一份放进第二份的“A”作为条件.

107

38. A、B、C 三位厨师烤某一种饼，烤坏的概率依次为 0.2，0.3，0.5. 若在他们工作的餐馆，在烤的这种饼中，A 占 50%，B 占 30%，C 占 20%. 问在烤坏的饼中，由 A 烤出的占多大比例？

39. 盒中有 3 枚硬币，其中第一枚两面都是正面，第二枚是均匀的，而第三枚是偏重的(它每次出现正面的机会占 75%). 现随机地从盒中取出一枚并抛出，结果出现正面，问它是第一枚硬币的概率是多少？

40. 监狱看守通知三个囚犯，在他们中要随机地选择一个处决，而把另两个释放. 囚犯 A 请求看守秘密告诉他，另外两个囚犯中谁将获得自由，并声称：“因为我已经知道他们中至少有一人获得自由，所以你泄露这点消息是无妨的. ”但是看守拒绝回答这个问题. 他对 A 说：“如果你知道了你的同伙中谁将获

释，那么你自己将被处决的概率就由$\frac{1}{3}$增加到$\frac{1}{2}$，因为你就成为剩下的两个囚犯中的一个了.”对于看守的上述理由，你是怎么想的?

41. 假定有10枚硬币，抛第i枚硬币出现正面的概率为$i/10(i=1,2,\cdots 10)$. 现随机地选择一枚硬币抛出，结果出现正面，问它是第5枚硬币的条件概率是多少?
42. 某年内，男性有车族给车上保险的概率为p_m，而女性为p_f，其中$p_f\neq p_m$. 保险客户中男性的比例为$\alpha(0<\alpha<1)$，随机选出一个保险客户，设A_i表示这个保险客户在第i年上保险的事件，试证：
$$P(A_2|A_1)>P(A_1)$$
对上述不等式给出一个直观的解释.
43. 箱中有5个白球与10个黑球. 现掷一个均匀的骰子，掷出几点就从箱中取出几个球. 问取出的全是白球的概率是多少? 若已知取出的全是白球，掷出的骰子是3点的条件概率是多少?
44. 两个外形相同的柜橱各有两个抽屉. A橱的每个抽屉各放一枚银币，B橱的一个抽屉放一枚银币、另一个抽屉放一枚金币. 现随机地选出一柜橱，打开一个抽屉，发现里面有一枚银币. 这个橱的另一个抽屉中也有一枚银币的概率是多少?
45. 假设有一种诊断癌症的试验，将它用于患癌症或不患癌症的人身上都有95%的准确率. 如果患癌症的人占4%，现对某人作这一试验，结果说明他有患癌症，试求他确实患此病的概率.
46. 某保险公司把保险人分为三类：“好的”、“一般的”与“坏的”. 统计资料表明，对于上述三类人而言，在一年内卷入某一次事故的概率依次为0.05，0.15，0.30. 如果“好的”保险人占保险人口的20%，“一般的”占保险人口的50%，“坏的”占保险人口的30%. 试问在固定的一年内出事故的人占多大比例? 如果某保险人在1997年没出事故，他是“好的”(“一般的”)概率是多少?
47. 某职员为找一份新工作希望她的上司提供一份推荐信. 她估计如果有一份好的推荐信就有80%的机会得到新工作，一般的推荐信有40%的机会得到新工作，差的推荐信有10%的机会得到新工作. 她又估计得到推荐信是好的、一般的、差的概率分别为0.7、0.2、0.1.

108

(a) 她有多大可能得到新工作?

(b) 已知她得到新工作，试问收到好的、一般的、差的推荐信各有多少可能?

(c) 已知她没有得到新工作，试问收到好的、一般的、差的推荐信各有多少可能?

48. 某高中毕业生很焦急地等待大学录取通知书，已知她被接受或拒绝，估计她下周每天收到通知书的条件概率如下：

天	P(信|接受)	P(信|拒绝)
星期一	0.15	0.05
星期二	0.20	0.10
星期三	0.25	0.10
星期四	0.15	0.15
星期五	0.10	0.20

她估计被接受的概率为0.6.

(a) 星期一收到信的概率是多少?

(b) 若星期一没有收到信，求星期三收到信的条件概率.

(c) 若星期三没有收到信，试求他被接受的条件概率.

(d) 若星期二收到信，求他被接受的条件概率.

(e) 若下周没有收到信，求他被接受的条件概率.

49. 某并联系统若至少一个部件运行则整个系统正常运行，考虑一个有n个部件的系统，每个部件运行的

概率为$\frac{1}{2}$，且是独立的. 已知系统运行，试求部件 1 运行的条件概率.

50. 如果你必须对下面(a)至(e)中所述的事件 E 与 F 构造数学模型，是否假定它们是独立的? 试述理由.
    (a) E 表示某商人的眼睛是蓝的这一事件，F 表示她的秘书的眼睛也是蓝的这一事件.
    (b) E 表示某教授拥有小汽车的事件，F 表示他的电话号码被列在电话簿上的事件.
    (c) E 表示某人身高不到 6 英尺[1]的事件，F 表示他的体重超过 200 磅[2]的事件.
    (d) E 表示某人生活在美国的事件，F 表示她生活在西半球的事件.
    (e) E 表示明天将下雨的事件，F 表示后天将下雨的事件.
51. 一个教室里有 4 个一年级男生，6 个一年级女生，6 个二年级男生. 为在随机选出一个学生时，性别与年级是相互独立的，教室里还应有多少个二年级女生?
52. 假定你不断地收集票券，收集到 m 个不同的类型. 另外每次只能收集一种新的类型，收集第 i 种类型的概率为 $p_i(i=1,\cdots,m)$. 若你已收集到第 n 种，试问它是一种新类型的概率是多少?
    提示:以这种票券的类型作为条件.

109

53. 股票价格的变化的一个简化模型是每天价格上升 1 个单位的概率为 p，下降 1 个单位的概率为 $1-p$. 假定每天价格的变化是独立的.
    (a) 2 天后股价不变的概率是多少?
    (b) 3 天后股价上升 1 个单位的概率是多少?
    (c) 已知 3 天后股价上升 1 个单位，第一天上升 1 个单位的条件概率是多少?
54. 假定我们归纳抛一枚均匀硬币的结果，但是抛偏重硬币时出现正面的概率为未知的 p，而不是$\frac{1}{2}$. 用下列步骤完成要求:
    1) 抛硬币.
    2) 再抛硬币.
    3) 如果两次都出现正面或反面，返回步骤 1.
    4) 以最后一次抛得的结果作为试验的结果.
    (a) 试证出现正面或反面是等可能的.
    (b) 用一个更简单的步骤完成最后两次抛硬币结果不一样，最后一次抛的结果是试验的结果.
55. 抛一枚硬币出现正面的概率为 p，试求前四次是(a) H,H,H,H;(b) T,H,H,H;(c) 在 H,H,H,H 之前 T,H,H,H 发生的概率.
    提示: 对(c)，考虑如何能得到 H,H,H,H?
56. 人眼睛的颜色是由一对基因决定的. 如果都是蓝眼睛的基因，则此人一定是蓝眼睛; 如果都是棕色眼睛的基因，则此人一定是棕色眼睛; 另外，如果有一个蓝眼睛的基因而另一个是棕色眼睛的基因，则此人是棕色眼睛.(因为我们认为棕色眼睛的基因比蓝色眼睛的基因有优势.)一个新生儿独立地从他父母亲处各获得一个眼睛基因，而且是等可能的. 假定史密斯及他的父母都是棕色眼睛，但是他的姐姐是蓝眼睛.
    (a) 史密斯有一个蓝眼睛基因的概率是多少?
    (b) 假定史密斯的妻子是蓝眼睛. 他们的第一个孩子是蓝眼睛的概率是多少?
    (c) 如果他的第一个孩子是棕色眼睛，则下一个孩子也是棕色眼睛的概率是多少?
57. 白化病的基因可记为 A 与 a，一个人从父母那里都得到基因 a，则他就是白化病患者. 如果有基因 A

---

[1] 英尺的单位符号为 ft，1 ft=0.304 8 m.——编辑注

[2] 磅的单位符号为 lb，1 lb=0.453 592 37 kg.——编辑注

与a，则这个人是正常的，但他可能把这个特性遗传到后代身上，称为携带者．假定一对正常夫妇有两个孩子，则恰有一个是白化病患者．假定一个没有白化病的人和一个白化病携带者结婚．

(a) 第一个孩子是白化病的概率是多少？

(b) 已知他们的第一个孩子不是白化病，试求第二个孩子是白化病的条件概率．

58. Barbara与Dianne去射靶．假定Barbara每次射中木质躲闪靶的概率为p_1，而Dianne的概率为p_2．假定两人同时射靶，如果靶被打倒了(说明被击中)，试求下列概率：

110

(a) 两个人都射到靶上；(b) Barbara射到靶上．

你的独立性假设是什么？

59. A、B两人决斗，决斗的规则是拿好枪，然后两人同时射击对方．假定一人被射中或两人都被射中则决斗结束，如果两人都射空了，则重复做．假定每次射击的结果是独立的，每次A射中B的概率为p_A，B射中A的概率为p_B．试求：

(a) A没被射中的概率．

(b) 双方都被射中的概率．

(c) 经过n次射击决斗结束的概率．

(d) 已知A没被射中，经过n次射击决斗结束的条件概率．

(e) 已知双方都被射中，经过n次射击决斗结束的条件概率．

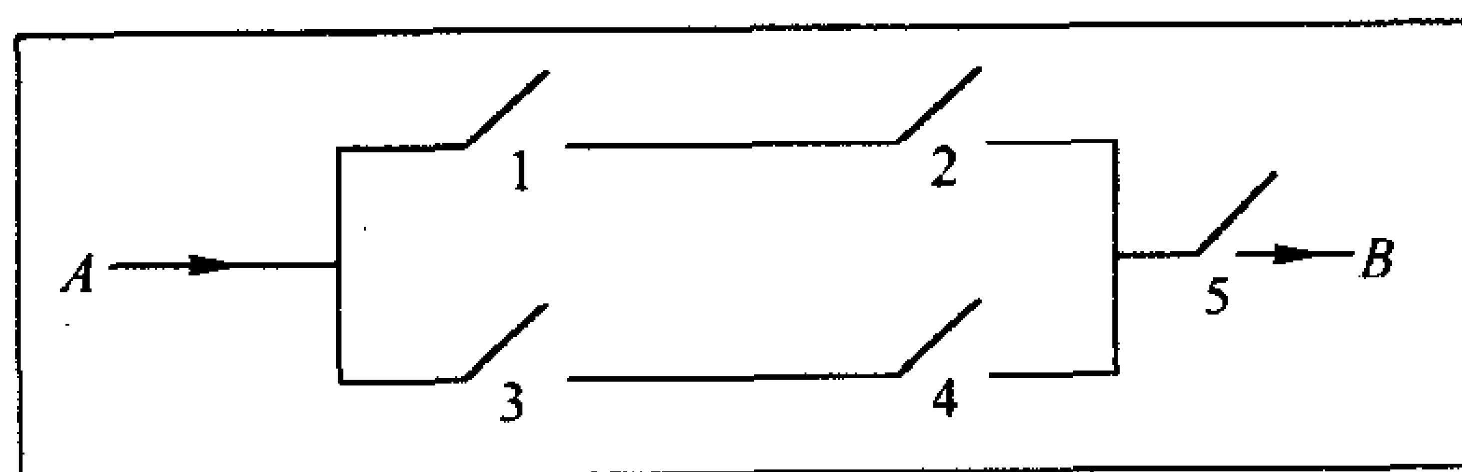
60. 在一个智力竞猜节目中，要求夫妻组队参加，回答"真假"问题．丈夫、妻子独立地给出正确答案的概率为p，下列哪种策略对这对夫妻来说比较好？

(a) 随机选出一人来回答问题．

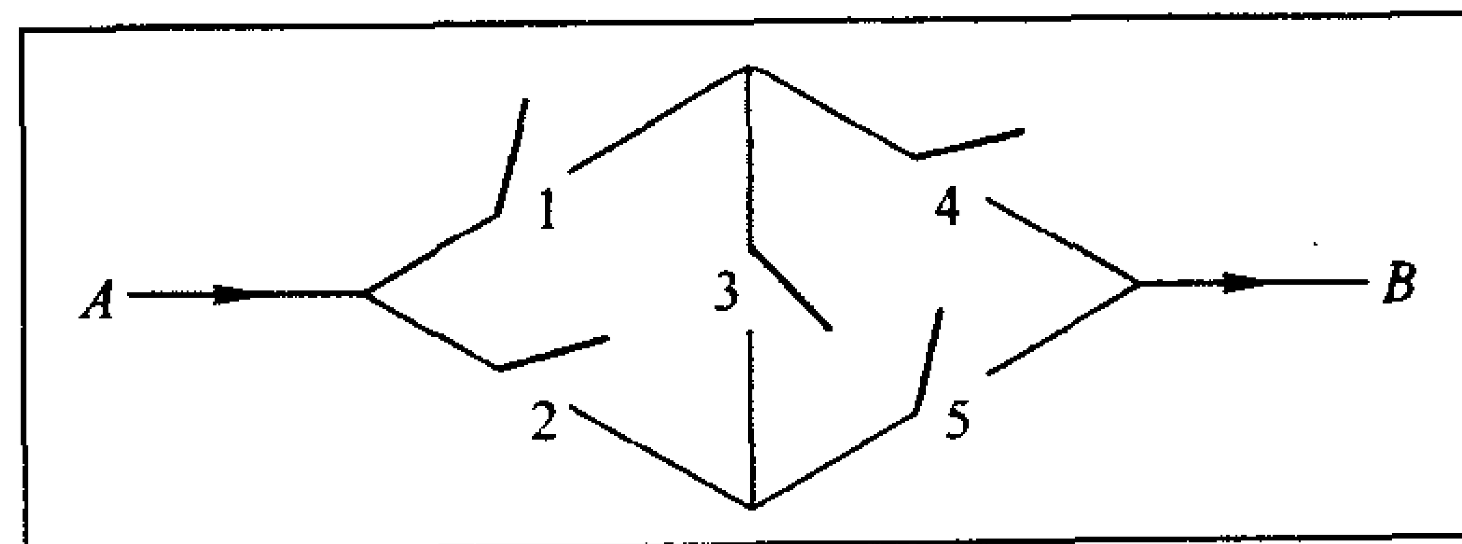
(b) 双方都考虑问题，或者两人的答案一致，则直接给出答案；或者两人的答案不一致，通过抛硬币来决定谁来回答问题．

61. 在60题中，若$p=0.6$，应该采用策略(b)，试求这对夫妇在(a)答案一致；(b)答案不一致的情况下给出正确答案的条件概率．

62. 在如下图所示的电路中，第i个继电器闭合的概率为$p_i, i=1,2,3,4,5$．如果所有继电器的功能是独立的，试对如下两个电路求出A与B之间是通路的概率．



a)



b)

提示：对于b)，以继电器3是否闭合作为条件．

63. 一个由n个部件组成的工程系统称为"k-n"系统($k \leqslant n$)，如果此系统工作当且仅当它的n个部件中至少k个部件工作，假设它的各部件工作是独立的．

111

(a) 若第i个部件工作的概率为$P_i, i=1,2,3,4$，求一个"2-4"系统工作的概率．

(b) 试对"3-5"系统求出上述概率．

(c) 假设一个"k-n"系统所有的 P_i 都等于 p(即 $P_i=p$, $i=1,2,\cdots,n$)，试对此系统求出上述概率.

64. 在题 62(a)中，已知 A 与 B 之间是通路，试求继电器 1 与 2 都闭合的条件概率.

65. 某生物体有 5 对基因(我们用前 5 个英文字母表示). 每个基因有两种形式(用大小写区别). 假定大写字母表示起决定作用的基因. 若某生物体有基因 Xx，则会显示 X 的特征. 例如，若 X 表示棕色眼睛，x 表示蓝眼睛，则有基因 XX 或 Xx 的人是棕色眼睛，有基因 xx 的人是蓝眼睛. 某生物体有典型特征称为显型，其遗传体征称为基因型. (两个生物体分别有遗传基因 aA, bB, cc, dD, ee 与 AA, BB, cc, DD, ee，基因型不同，但显型相同.)在两个生物体的配对中，每个生物体随机地贡献一种类型的基因对. 假定一个生物体的 5 对基因是独立的，对配对的贡献也是独立的. 在一次配对中，两个生物体分别有遗传基因 aA, bB, cC, dD, eE 与 aa, bB, cc, Dd, ee. 试求其后代(i) 显型 (ii) 基因型类似于下列情况的概率：(a) 第一个生物体；(b) 第二个生物体；(c) 任何一个生物体；(d) 两个生物体都不是.

66. 皇后携带血友病基因的机会为 50-50. 如果她是一个携带者，则每位王子有 50-50 的机会是血友病. 如果皇后有三个王子都没有此病，试求皇后是携带者的概率. 如果还有第四个王子，该王子患此病的概率是多少?

67. 1982 年 9 月 30 日早上，美国西部联赛垒球比赛前三名的结果如下表：

球 队	赢	输
亚特兰大勇士队	87	72
旧金山巨人队	86	73
洛杉矶道奇队	86	73

每个球队接下来要进行 3 场比赛，旧金山巨人队的 3 场比赛对手是洛杉矶道奇队，亚特兰大勇士队的 3 场比赛对手是圣迭哥教士队. 假定每次比赛的结果是独立的，每场比赛可以由任意一队获胜. 试求每队赢得决赛的概率是多少? 如果两个队并列第一，则要进行加时赛，每队获胜是等可能的.

68. 某镇委会有 7 个成员，其中有 3 个筹委会成员. 立法的一个新方案先交给筹委会的 3 个成员，若至少 2 个成员同意此方案，则交给镇委会. 如果镇委会中至少 4 人同意，则通过此方案. 每个成员接受此方案是独立的，概率为 p. 试问给定筹委会一位成员的投票起决定作用，如果她投反对票，则此方案没有通过的概率是多少? 镇委会(不包括筹委会成员)的对应概率又是多少? [112]

69. 假设一对夫妇生男孩与女孩的概率是等可能的，且与他们的其他孩子的性别无关. 对于有 5 个孩子的夫妇，试求如下事件的概率.
   (a) 所有孩子性别相同.
   (b) 3 个大的是男孩，另 2 个是女孩.
   (c) 正好 3 个男孩.
   (d) 2 个大的是女孩.
   (e) 至少 1 个女孩.

70. A、B 掷一对均匀的骰子，直到 A 的两个骰子的点数之和为 9 或者 B 的两个骰子的点数之和为 6 为止. 假定 A 先掷，试求 A 掷最后一次的概率.

71. 某村有一传统，长子有责任赡养父母. 近几年来，妇女不想承担责任，因此不愿嫁给长子.
   (a) 如果该村每个家庭都有两个孩子，试问所有的男孩是长子占多大比例?
   (b) 如果每个家庭有三个孩子，试问所有的男孩是长子占多大比例?
   假定每个孩子的性别是独立的.

72. 设 E 与 F 是某一试验的两个互不相容事件. 试证：若将此试验独立重复进行，那么 E 将在 F 之前发生的概率为 $P(E)/[P(E)+P(F)]$.

73. A、B进行一系列比赛，每场比赛是独立的，A每场获胜的概率为p，B相应的概率为$1-p$. 当其中一人获胜的次数是另一人的两倍多时，停止比赛. 获胜次数多的人将是最后的胜利者.
(a) 试求一共进行4场比赛的概率.
(b) 试求A是胜利者的概率.

74. 接连掷一对均匀的骰子，问在得到6次偶数点之前，掷得两次7点的概率是多少?

75. 选手的技艺相同，且在每一场比赛中，参加比赛的两名选手每人获胜的概率均为$\frac{1}{2}$. 将2^n名选手随机地配成两人一对独立进行比赛；然后将2^{n-1}名获胜者再随机地配对，如此继续下去，直到只剩下一名优胜者. 考虑两个指定的选手A、B，并定义事件$A_i(i\leqslant n)$与事件E如下：
A_i：A正好参加了i场比赛.
E：A与B曾对阵.
(a) 求$P(A_i), i=1,\cdots,n$.
(b) 求$P(E)$.
(c) 设$P_n=P(E)$，试证
$$P_n=\frac{1}{2^n-1}+\frac{2^n-2}{2^n-1}\left(\frac{1}{2}\right)^2 P_{n-1}$$
并用来验证(b)中的答案.
提示：利用事件$A_i(i=1,2,\cdots,n)$是否发生为条件计算$P(E)$. 用代数等式
$$\sum_{i=1}^{n-1} ix^{i-1}=\frac{1-nx^{n-1}+(n-1)x^n}{(1-x)^2}$$
简化答案. 用另外一种方法也可以解此问题，注意共有2^n-1场比赛.
(d) 说明为什么有2^n-1场比赛.
用数字标出各场比赛，B_i表示A与B在第$i(i=1,2,\cdots,2^n-1)$场比赛.
(e) 求$P(B_i)$.
(f) 试用$P(B_i)$求$P(E)$.

113

76. 某股票市场投资者有一份现值为25的股票，如果该股票降到10或升到40，她决定抛出. 若股票的每次变化以概率0.55升1，以概率0.45降1. 每次变化是独立的，试求她赔了的概率.

77. A、B抛一枚均匀的硬币，A先连续地抛直到出现反面为止，这时B接着抛直到出现一个反面. 然后A再接着抛，如此继续下去. 设A抛硬币出现正面的概率为P_1，B抛出正面的概率为P_2. 若约定先得到如下结果者获胜：
(a) 2个连续的正面.
(b) 正面累计数为2.
(c) 3个连续的正面.
(d) 正面累计数为3.
在每一种情形下，求A获胜的概率.

78. 骰子A有4个红面与2个白面，而骰子B有4个白面与2个红面. 抛某均匀的硬币一次，若出现正面，则用骰子A进行比赛，若出现反面，则用骰子B进行比赛.
(a) 试证：每一次掷骰子时，红面出现的概率为$\frac{1}{2}$.
(b) 如果前两次掷骰子的结果是红面，求第三次也掷出红面的概率是多少?
(c) 如果前两次掷骰子出现红面，求所用的是骰子A的概率是多少?

79. 箱中有12个球，其中4个白球，A、B、C三个人接连地从箱中取球，A先取，然后B，然后C，然后A，如此继续下去. 规定先取出一个白球者获胜. 如果(a)每个球取出后都放回箱中，(b)取出的球不

放回箱中，分别求出 A、B、C 获胜的概率.

80. 如果有三个不同的箱子，每箱各有 12 个球，其中 4 个白球. A、B、C 各从自己的箱中取球，上题的答案又是多少?

81. 设 $S=\{1,2,\cdots,n\}$，假定 A 与 B 独立地等可能地是 S 的 2^n 个子集合中的一个(包含零集与 S).

(a) 试证：

$$P\{A\subset B\}=\left(\frac{3}{4}\right)^n$$

提示：设 $N(B)$表示 B 中的元素的个数，运用

$$P\{A\subset B\}=\sum_{i=0}^{n}P\{A\subset B\mid N(B)=i\}P\{N(B)=i\}$$

(b) 证明 $P\{AB=\varnothing\}=\left(\frac{3}{4}\right)^n$.

82. 在例 5d 中，已知前 n 次试验的所有结果都是正面，被选到的是第 i 枚硬币的条件概率是多少?

83. 在拉普拉斯继承准则(例 5d)中，连续抛的结果是不是独立的? 试说明理由.

84. 由编号为 1,2,3 的三名法官组成一陪审团. 若某人被审讯后，至少两名法官投有罪票，则判决此人有罪. 假设对一名事实上有罪的被告，每个法官独立地投有罪的概率皆为 0.7，而对事实上无罪的被告，这个概率下降到 0.2. 如果 70%的被告是事实上有罪的，试对如下各条件求出 3 号法官投有罪票的条件概率. [114]

(a) 1 号与 2 号法官都投有罪票.

(b) 在 1 号与 2 号法官所投的票中，一张有罪，另一张无罪.

(c) 1 号与 2 号法官都投无罪票.

以 $E_i(i=1,2,3)$表示 i 号法官投一张有罪票的事件. 这些事件是否独立? 是否条件独立? 试说明理由.

85. 假定进行 n 次独立试验，每次分别以概率 p_0,p_1,p_2 出现结果 0,1,2, $\sum_{i=0}^{2}p_i=1$. 试求至少一次出现结果 1 和 2 的概率.

理论练习

1. 试证：若 $P(A)>0$，则 $P(AB|A)\geqslant P(AB|A\cup B)$.

2. 若 $A\subset B$，尽可能简洁地表示下列概率：

$$P(A|B),\ P(A|B^c),\ P(B|A),\ P(B|A^c)$$

3. 有 m 个家庭组成师生社团，n_i 个人中有 i 个孩子，$i=1,\cdots,k$，$\sum_{i=1}^{k}n_i=m$. 可以通过下列两种方法选出一个孩子：

(a) 从 m 个家庭任意选出一个家庭，从那个家庭再任意选一个孩子.

(b) 从 $\sum_{i=1}^{k}in_i$ 个孩子中任意选一个.

证明方法 1 比方法 2 容易选出头胎孩子.

提示：为证明结论，只需证明下式：

$$\sum_{i=1}^{k}in_i\sum_{j=1}^{k}\frac{n_j}{j}\geqslant\sum_{i=1}^{k}n_i\sum_{j=1}^{k}n_j$$

要证明上式，先对各和作乘法，然后证明对所有对 i，j，左边表达式中项 n_in_j 的系数大于右边表达式中项 n_in_j 的系数.

4. 有 n 个盒子，一个球在第 i 个盒中的概率为 P_i，经检查这个盒子能发现球的概率为 α_i，试证明检查了第

i 个盒子没找到球，而它事实上在第 j 个盒中的条件概率为

$$\begin{cases} \dfrac{P_j}{1-\alpha_i P_i} & 若\ j\neq i \\ \dfrac{(1-\alpha_i)P_i}{1-\alpha_i P_i} & 若\ j=i \end{cases}$$

5. 事件 F 称为不利于事件 E，记作 $F\searrow E$，如果

115

$$P(E|F)\leqslant P(E)$$

试证明如下论断或给出反例

(a) 若 $F\searrow E$，则 $E\searrow F$.

(b) 若 $F\searrow E$ 且 $E\searrow G$，则 $F\searrow G$.

(c) 若 $F\searrow E$ 且 $G\searrow E$，则 $FG\searrow E$.

类似可定义 F 有利于 E，记作 $F\nearrow E$，如果 $P(E|F)\geqslant P(E)$. 将$\searrow$改为$\nearrow$，重做(a)、(b)和(c).

6. 试证：若 $E_1,E_2,\cdots,E_n$ 是独立事件，则

$$P(E_1\cup E_2\cup\cdots\cup E_n)=1-\prod_{i=1}^{n}[1-P(E_i)]$$

7. (a) 箱中有 n 个白球与 m 个黑球，球被一个一个地取出直到箱中只剩下同样颜色的球为止. 试证其余全是白球的概率为 $n/(n+m)$.

   提示：想象一下试验继续进行直到所有球被取出，并考虑最后一个取出的球.

   (b) 池塘里有 3 种不同类型的鱼，我们称之为红(R)、蓝(B)、绿(G). 共有 r 条红鱼，b 条蓝鱼，g 条绿鱼，从池塘中随机依次取出鱼(也就是说，每次等可能地从池塘的剩余鱼中取出). 试求池塘中红鱼最先取完的概率.

   提示：运用 $P\{R\}=P\{RBG\}+P\{RGB\}$，先把最后取完的类型的鱼作为条件来计算右式的概率.

8. 抛两枚均匀的硬币，A 表示抛第一次出现正面的事件，B 表示抛第二次出现正面的事件，C 表示两次抛硬币出现同正或同反的事件. 试证明 A，B，C 是两两独立的，即 A 与 B 是相互独立的，A 与 C 是相互独立的，B 与 C 也是相互独立的. 但 A，B，C 不是相互独立的.

9. 由 n 个人组成的聚会，若每个人的生日是 365 天中任意一天，并且这些生日是独立的. 令 $A_{i,j}$表示 i 和 j 两个人生日相同的事件，其中 $i\neq j$. 试证明这些事件是两两独立的，即 $A_{i,j}$ 与 $A_{r,s}$是独立的，但 $\binom{n}{2}$个事件 $A_{i,j}(i\neq j)$不独立.

10. 一枚均匀的硬币连续抛 n 次，出现正面的概率为 p，问 n 要有多大才能使至少出现一个正面的概率至少为$\frac{1}{2}$?

11. 设 $0\leqslant a_i\leqslant 1$，$i=1,2,\cdots$，证明

$$\sum_{i=1}^{\infty}\left[a_i\prod_{j=1}^{i-1}(1-a_j)\right]+\prod_{i=1}^{\infty}(1-a_i)=1$$

12. 抛某硬币一次出现正面的概率为 p. 如果 A 连续地抛这枚硬币直到出现反面为止，此后 B 接着抛，再连续地抛到出现反面为止，此后 A 再连续抛，如此继续下去. 以 $P_{n,m}$表示在 B 积累了 m 个正面之前 A 积累了 n 个正面的概率. 试证：

$$P_{n,m}=pP_{n-1,m}+(1-p)(1-P_{m,n})$$

116

*13. 假如你与一个无限富有的对手赌钱，每赌一局你赢 1 元的概率为 p，输 1 元的概率为 $1-p$，试证你输光的概率为

$$\begin{cases} 1 & 若\ p\leqslant\dfrac{1}{2} \\ (q/p)^i & 若\ p>\dfrac{1}{2} \end{cases}$$

其中 $q=1-p$，i 是你的赌本.

14. 一个独立重复试验，每次成功的概率为 p，现连续进行下去直到获得 r 次成功为止. 证明需进行正好 n 次试验的概率为

$$\binom{n-1}{r-1}p^r(1-p)^{n-r}$$

试运用这一结果解分赌注问题(例 4i).

提示:为使 n 次试验中有 r 次成功,在前 $n-1$ 次试验中必须成功多少次?

15. 若独立重复试验每次成功的概率为 p，失败的概率为 $1-p$，则称之为伯努利试验. 令 P_n 表示 n 次伯努利试验有偶数次成功的概率(认为 0 是一个偶数). 试证:

$$P_n=p(1-P_{n-1})+(1-p)P_{n-1} \quad n\geqslant 1$$

由此(利用归纳法)证明

$$P_n=\frac{1+(1-2p)^n}{2}$$

16. 令 Q_n 表示抛一枚均匀硬币 n 次，而不出现连续 3 次正面的概率. 试证:

$$Q_n=\frac{1}{2}Q_{n-1}+\frac{1}{4}Q_{n-2}+\frac{1}{8}Q_{n-3}$$
$$Q_0=Q_1=Q_2=1$$

计算 Q_8.

提示: 以第一次出现反面作为条件.

17. 考虑赌徒输光问题. 假定 A、B 商量好至少赌 n 次，以 $P_{n,i}$表示 A 开始有 i 元、B 有 $N-i$ 元而 A 赢所有钱的概率. 导出用 $P_{n-1,i+1}$与 $P_{n-1,i-1}$表示 $P_{n,i}$的等式，并计算 $N=5$ 时的 $P_{7,3}$.

18. 两个箱中都既有白球又有黑球，从第一、第二箱中任取一球是白球的概率分别为 p 和 p'. 现接连从箱中有放回地取球如下: 头一个球是从第一箱中抽取的概率为 α，从第二箱抽取的概率为 $1-\alpha$. 以后的抽取按照下面的规则: 若前一次取出的球是白的，则下一次就从同一箱中抽球. 若前一次取出的球是黑的，则下一次就从另一箱中抽球. 以 α_n 表示第 n 个球是从第一箱中抽取的概率，试证:

$$\alpha_{n+1}=\alpha_n(p+p'-1)+1-p' \qquad n\geqslant 1$$

并用上式证明

$$\alpha_n=\frac{1-p'}{2-p-p'}+\left(\alpha-\frac{1-p'}{2-p-p'}\right)(p+p'-1)^{n-1}$$

117

设 P_n 表示第 n 次抽取的是白球的概率. 试求 P_n，并计算 $\lim\limits_{n\to\infty}\alpha_n$ 与 $\lim\limits_{n\to\infty}P_n$.

19. **投票问题**. 在一次选举中，候选人 A 收到 n 张选票，候选人 B 收到 m 张选票，其中 $n>m$. 假定选票的所有 $(n+m)!/n!\,m!$ 种排序是等可能的，令 $P_{n,m}$表示 A 的选票在选举中领先的概率.

(a) 计算 $P_{2,1}$，$P_{3,1}$，$P_{3,2}$，$P_{4,1}$，$P_{4,2}$，$P_{4,3}$.

(b) 计算 $P_{n,1}$，$P_{n,2}$.

(c) 基于(a)与(b)，猜测 $P_{n,m}$的值.

(d) 以谁收到最后一张选票作为条件，根据 $P_{n-1,m}$与 $P_{n,m-1}$得到 $P_{n,m}$的递归式.

(e) 用(d)验证(c)中对 $n+m$ 用归纳法的猜测.

20. 作为天气预报的一个简化模型，假设明天的天气(干燥或潮湿)与今天相同的概率为 p. 如果 1 月 1 日是干燥，试证接下来连续 n 天都是干燥的概率 P_n 满足

$$P_n=(2p-1)P_{n-1}+(1-p) \qquad n\geqslant 1$$
$$P_0=1$$

证明

$$P_n=\frac{1}{2}+\frac{1}{2}(2p-1)^n \quad n\geqslant 0$$

21. 袋中有 a 个白球与 b 个黑球，球按如下规则从袋中取出：
 (a) 随机地取出一球然后丢掉.
 (b) 如果再取出的球与前一个颜色不同就把它放回袋中，再从(a)重新开始. 如果再取出的球与前一个颜色相同就把它丢掉，再从(b)开始.
 换言之，球被一个一个地取出来、观察、丢弃，直到出现不同的颜色为止，此时把这最后一球放回袋中，过程重新开始. 令 $P_{a,b}$表示最后留在袋中的是一个白球的概率，试证：
$$P_{a,b}=\frac{1}{2}$$
 提示：对 $k\equiv a+b$ 用归纳法.

*22. 某循环赛共有 n 个选手参加，其中$\binom{n}{2}$对选手需互相比赛，每次比赛总有输赢的结果. 对某固定的整数 k，$k<n$，我们感兴趣的问题是对任意 k 个选手，循环赛的结果是总有一个选手需与他们进行比赛这是否可能. 试证：若
$$\binom{n}{k}\left[1-\left(\frac{1}{2}\right)^k\right]^{n-k}<1$$
则这样的结果是可能的.

提示：假定这些比赛结果是独立的，每个选手获胜是等可能的. 对 k 个选手的$\binom{n}{k}$个组合，设 B_i 表示第 i 个集中所有 k 个选手没有选手比赛的事件，然后运用布尔不等式求 $P\left(\bigcup_i B_i\right)$.

118

23. 直接证明
$$P(E|F)=P(E|FG)P(G|F)+P(E|FG^c)P(G^c|F)$$
24. 证明式(5-11)与式(5-12)的等价性.
25. 将条件独立性的定义推广到多于两个事件的情形.
26. 证明或给出反例：若 E_1 与 E_2 独立，则 E_1 与 E_2 对于 F 条件独立.
27. 在拉普拉斯继承准则(例 5d)中，证明：若前 n 次抛硬币皆出现正面，则接下来 m 次仍然全出现正面的条件概率为$(n+1)/(n+m+1)$.
28. 在拉普拉斯继承准则中，假设前 n 次抛出 r 个正面与 $n-r$ 个反面，试证第$(n+1)$次抛出一个正面的概率为$(r+1)/(n+2)$. 为此，你必须证明并运用等式
$$\int_0^1 y^n(1-y)^m\mathrm{d}y=\frac{n!m!}{(n+m+1)!}$$
 提示：为证这一等式，令 $C(n,m)=\int_0^1 y^n(1-y)^m\mathrm{d}y$. 用分部积分法可得
$$C(n,m)=\frac{m}{n+1}C(n+1,m-1)$$
 从 $C(n,0)=1/(n+1)$出发，对 m 作归纳法即可证明等式.
29. 假定你的一个不懂数学但有哲学头脑的朋友声称拉普拉斯继承准则必定是错误的，理由是它可以导出荒谬的结论. “例如，”他说，“如果一个孩子 10 岁，按这个原则，由于他已经活了 10 年，那么再活一年的概率为$\frac{11}{12}$. 另一方面，如果这个孩子有位 80 岁的祖父，那么由拉普拉斯准则，祖父再活一年的概率却是$\frac{81}{82}$. 但这是荒谬的，显然，孩子比祖父更容易再活一年. ”你怎么回答这位朋友?

自测题与练习

1. 在桥牌游戏中，西手中无 A，问他的同伴手中(a)无 A，(b)有 2 个或更多 A 的概率是多少？(c)若西手中正好有一个 A，则他的同伴手中无 A 或者有 2 个或更多 A 的概率各是多少？
2. 新车电池能跑 10 000 英里[⊖]的概率为 0.8，能跑 20 000 英里的概率为 0.4，超过 30 000 英里的概率为 0.1. 若新车跑完 10 000 英里仍能工作，问下列概率是多少？
   (a) 电池全部寿命超过 20 000 英里.
   (b) 电池额外寿命超过 20 000 英里.
3. 有 20 个球，其中 10 个白球，10 个黑球，放到两个箱中，试问如何放置才能使随机选的箱中随机取到白球的概率最大？
4. A 箱中有 2 个白球与 1 个黑球，而 B 箱中有 1 个白球与 5 个黑球，随机从 A 箱中抽出一球放入 B 箱，然后从 B 箱取出一球，发现是个白球. 问从 A 箱转移到 B 箱的那个球也是白球的概率为多少？

119

5. 箱中有 b 个黑球与 r 个红球，随机地取出一球，然后再添上 c 个与它同样颜色的球一起放回箱中. 现再取出一球，试证：若第二个球是红球，则第一个球是黑球的概率为 $b/(b+r+c)$.
6. 你的一位朋友随机从一副 52 张扑克牌中无放回地抽取两张，在下列每种情形下，试求两张都是 A 的条件概率.
   (a) 你问你的朋友是否其中一张是黑桃 A，他确定是.
   (b) 你问你的朋友是否第一张为 A，他确定是.
   (c) 你问你的朋友是否第二张为 A，他确定是.
   (d) 你问你的朋友是否其中一张是 A，他确定是.
7. 试证：

$$\frac{P(H|E)}{P(G|E)}=\frac{P(H)P(E|H)}{P(G)P(E|G)}$$

   假定观察到新证据前假设 H 为真的可能性是假设 G 的 3 倍，新证据在 G 下为真的可能性是在 H 下为真的两倍. 试问在观察到新条件下，哪一个假设更可能发生？
8. 在你外出度假时，你托邻居帮你浇快要凋谢的花. 不浇水花凋谢的概率为 0.8，浇水花仍会凋谢的概率为 0.15，你有 90%的把握确信邻居会记着帮你浇花.
   (a) 试问在你回来时，花活着的概率是多少？
   (b) 如果花凋谢了，试求你的邻居忘记浇花的概率.
9. 某种类型的老鼠，黑色控制棕色. 假定有黑色父母的某黑色老鼠有一个棕色同胞.
   (a) 试问这只老鼠是纯黑(不是一个黑基因与一个棕基因的混合)的概率是多少？
   (b) 假定某黑老鼠与一只棕老鼠交配，则所有的 5 个后代都是黑老鼠. 试问这只老鼠是纯黑的概率是多少？
10. (a) 在习题 62 中，若以继电器 1 是否闭合为条件，试求电流从 A 流到 B 的条件概率.
    (b) 若电流能从 A 流到 B，试求继电器 3 闭合的条件概率.
11. 在习题 63 的 k-n 系统中，假定每个部件独立，以概率 $\frac{1}{2}$ 正常工作. 试求已知整个系统正常工作，部件 1 在工作的条件概率.
    (a) $k=1,n=2$. (b) $k=2,n=3$.

[⊖] 英里的单位符号为 mile，1 mile=1 609.344 m. ——编辑注

12. 为在轮盘赌中获胜，琼斯设计了一个赌博方案：只有当轮盘在前10次旋转都停在黑数字上的时候，他才把赌注下在红数字上．他的理由是，接连11次旋转都停在黑数字上的概率十分小，因此他获胜的机会很大．对于这个方案，你是怎么想的？
13. 三个人同时抛硬币，A，B，C抛出正面的概率各为P_1，P_2，P_3．如果某个人得到了与其余两人不同的结果，则他就是优胜者．如果没出现优胜者，就继续抛，直到出现优胜者为止．问A将成为优胜者的概率是多少？

120

14. 假定独立试验有n个可能结果，结果i出现的概率为p_i，$i=1,\cdots,n$，$\sum_{i=1}^{n}p_i=1$．若观察两个独立试验，试问第二个试验结果大于第一个试验结果的概率是多少？
15. 若A有$n+1$枚均匀硬币，B有n枚均匀，他们把硬币一一抛出．证明A得到的正面比B多的概率是$\frac{1}{2}$．
提示：以他们都抛出n枚硬币后的情况作为条件，这时，比较他们所得的正面的多少(有三种可能)．
16. 证明或给出反例：
(a) 若E与F独立，E与G独立，则E与$F\cup G$独立．
(b) 若E与F独立，E与G独立，且$FG=\varnothing$，则E与$F\cup G$独立．
(c) 若E与F独立，F与G独立，E与FG独立，则G与EF独立．
17. 设A与B都有正的概率．试说明下列事件是(i)一定为真，(ii)一定为假，(iii)可能为真．
(a) 如果事件A与B互不相容，则它们是独立的．
(b) 如果事件A与B是独立的，则它们互不相容．
(c) 若$P(A)=P(B)=0.6$，则它们互不相容．
(d) 若$P(A)=P(B)=0.6$，则它们是独立的．
18. 试比较下列事件发生的可能性(从最可能到最不可能)．
(a) 一枚均匀的硬币出现正面．
(b) 3次独立试验，每次成功的概率为0.8，3次都成功．
(c) 7次独立试验，每次成功的概率为0.9，7次都成功．
19. 当地有两个工厂生产收音机，A厂生产次品的概率为0.05，B厂生产次品的概率为0.01．假定你买了同一厂生产的两个收音机，但不清楚是哪一个厂的．若你发现第一个收音机是次品，试问第二个收音机也是次品的条件概率是多少？
20. 试证：若$P(A|B)=1$，则$P(B^c|A^c)=1$．
21. 箱中最初有1个红球与1个蓝球，每次从箱中随机无放回地取出一个球，再放入2个同色的球．(例如，取出1个红球，等到第二次取球时，箱中有2个红球与1个蓝球．)运用数学归纳法证明经过n次取球后，箱中有i个红球的概率为$\frac{1}{n+1}$，$1\leqslant i\leqslant n+1$．
22. $2n$张扑克牌(其中2张A)分给两个人，每人有n张．如果分到A要说明．已知第一人有A，求第二人没有A的条件概率，若(a) $n=2$，(b) $n=10$，(c) $n=100$．若n趋于无穷，这个概率的极限是多少？为什么？

121

第4章 随机变量

4.1 随机变量

经常发生这种情况，进行一次试验时，我们主要感兴趣的是试验结果的某个函数，而不是试验结果本身. 例如，掷两个骰子，通常感兴趣的是它们的点数之和而不管这两个骰子各是几点. 这就是说，比如感兴趣的是点数之和等于 7，而不在乎实际结果是(1,6)，(2,5)，(3,4)，(4,3)，(5,2)还是(6,1). 又如在掷硬币时，可能只对掷出正面的总数感兴趣，而不关心结果序列中正面、反面如何排列. 感兴趣的这些量，或更严格地说，这些定义于样本空间上的实值函数称为随机变量.

由于随机变量的值由试验结果所决定，因此可以研究随机变量可能值的概率分布.

例 1a 假定掷 3 枚均匀的硬币，以 Y 表示正面出现的次数，那么 Y 是一个随机变量，它取值为 0，1，2，3 的概率分别为

$$P\{Y=0\}=P\{(T,T,T)\}=\frac{1}{8}$$

$$P\{Y=1\}=P\{(T,T,H),(T,H,T),(H,T,T)\}=\frac{3}{8}$$

$$P\{Y=2\}=P\{(T,H,H),(H,T,H),(H,H,T)\}=\frac{3}{8}$$

$$P\{Y=3\}=P\{(H,H,H)\}=\frac{1}{8}$$

因为 Y 必定取 0～3 的某一整数，所以有

$$1=P\left(\bigcup_{i=0}^{3}\{Y=i\}\right)=\sum_{i=0}^{3}P\{Y=i\}$$

当然这个公式和以前所学的一致. ■

例 1b 箱中有编号为 1 至 20 的 20 个球，现无放回地随机取出 3 个球. 若打赌至少有 1 个被取出的球编号不小于 17 获胜，问获胜的概率是多少？

122

解 令 X 表示所取出球的最大编号，那么 X 是在 $3,4,\cdots,20$ 中取值的随机变量. 另外，若假定 $\binom{20}{3}$ 种可能取法中每一种取法的出现是等可能的，则有

$$P\{X=i\}=\frac{\binom{i-1}{2}}{\binom{20}{3}},i=3,4,\cdots,20 \tag{1-1}$$

式(1-1)之所以成立，是因为在所有 $\binom{20}{3}$ 种取法中，使事件 $\{X=i\}$ 发生的取法正是取到一个编号为 i 的球，而另两个球的编号在 1 至 $i-1$ 之间. 显然这种取法有 $\binom{1}{1}\binom{i-1}{2}$ 种，所以就得到概率的表达式(1-1). 由这个等式得到

$$P\{X=20\}=\frac{\binom{19}{2}}{\binom{20}{3}}=\frac{3}{20}=0.150$$

$$P\{X=19\}=\frac{\binom{18}{2}}{\binom{20}{3}}=\frac{51}{380}\approx 0.134$$

$$P\{X=18\}=\frac{\binom{17}{2}}{\binom{20}{3}}=\frac{34}{285}\approx 0.119$$

$$P\{X=17\}=\frac{\binom{16}{2}}{\binom{20}{3}}=\frac{2}{19}\approx 0.105$$

由于事件$\{X\geqslant 17\}$是不相交事件$\{X=i\}(i=17,18,19,20)$的并，故获胜的概率为

$$P\{X\geqslant 17\}\approx 0.105+0.119+0.134+0.150=0.508$$ ■

123

例 1c 掷一枚硬币出现正面的概率为 p，现重复地进行这一试验，直到出现一个正面或者掷了 n 次为止. 如果令 X 表示掷硬币的次数，那么 X 是一个随机变量，它取值 $1,2,\cdots,n$ 的概率分别为

$$P\{X=1\}=P\{H\}=p$$

$$P\{X=2\}=P\{(T,H)\}=(1-p)p$$

$$P\{X=3\}=P\{(T,T,H)\}=(1-p)^2p$$

$$\vdots$$

$$P\{X=n-1\}=P\{(\underbrace{T,T,\cdots,T}_{n-2},H),=(1-p)^{n-2}p$$

$$P\{X=n\}=P\{(\underbrace{T,T,\cdots,T}_{n-1},T),(\underbrace{T,T,\cdots,T}_{n-1},H)\}=(1-p)^{n-1}$$

作为一个验证，注意到

$$P\left(\bigcup_{i=1}^{n}\{X=i\}\right)=\sum_{i=1}^{n}P\{X=i\}=\sum_{i=1}^{n-1}p(1-p)^{i-1}+(1-p)^{n-1}$$

$$=p\left[\frac{1-(1-p)^{n-1}}{1-(1-p)}\right]+(1-p)^{n-1}=1-(1-p)^{n-1}+(1-p)^{n-1}=1$$ ■

例 1d 从装有 3 个白球、3 个红球与 5 个黑球的箱中随机地取出 3 个球. 假定每取出 1 个白球赢 1 美元，每取出 1 个红球输 1 美元. 现以 X 表示在这一试验中赢钱的总数，则 X 为一随机变量，它取值 $0,\pm1,\pm2,\pm3$ 的概率分别为

$$P\{X=0\}=\frac{\binom{5}{3}+\binom{3}{1}\binom{3}{1}\binom{5}{1}}{\binom{11}{3}}=\frac{55}{165}$$

$$P\{X=1\}=P\{X=-1\}=\frac{\binom{3}{1}\binom{5}{2}+\binom{3}{2}\binom{3}{1}}{\binom{11}{3}}=\frac{39}{165}$$

124

$$P\{X=2\}=P\{X=-2\}=\frac{\binom{3}{2}\binom{5}{1}}{\binom{11}{3}}=\frac{15}{165}$$

$$P\{X=3\}=P\{X=-3\}=\frac{\binom{3}{3}}{\binom{11}{3}}=\frac{1}{165}$$

这些概率是这样得到的，例如，为使 $X=0$，则要么取出的 3 个全为黑球，要么三种颜色各取出 1 个．类似地，事件$\{X=1\}$发生意味着要么取出 1 个白球与 2 个黑球，要么取出 2 个白球与 1 个红球．作为一个验证，注意到

$$\sum_{i=0}^{3}P\{X=i\}+\sum_{i=1}^{3}P\{X=-i\}=\frac{55+39+15+1+39+15+1}{165}=1$$

则赢钱的概率为

$$\sum_{i=1}^{3}P\{X=i\}=\frac{55}{165}=\frac{1}{3}$$ ■

例 1e 设有 N 种不同类型的票券，每次只能获得一种票券，且独立于前面的选择，即等可能地获得 N 种票券中的任何一种．T 表示为获得整套票券所收集的票券数，不直接计算 $P\{T=n\}$，而是先计算 T 大于 n 的概率．固定 n，定义事件 $A_1,A_2,\cdots,A_N$ 如下：A_j 表示前 n 种票券没有第 j 型的事件，其中 $j=1,2,\cdots,N$．因此

$$\begin{aligned}P\{T>n\}=P\Big(\bigcup_{j=1}^{N}A_j\Big)&=\sum_{j}P(A_j)-\sum_{j_1<j_2}\sum P(A_{j_1}A_{j_2})+\cdots\\&+(-1)^{k+1}\sum_{j_1<j_2<\cdots<j_k}\sum\sum P(A_{j_1}A_{j_2}\cdots A_{j_k})+\cdots+(-1)^{N+1}P(A_1A_2\cdots A_N)\end{aligned}$$

如果前 n 种票券没有第 j 型则事件 A_j 发生．因为票券没有第 j 型的概率为$(N-1)/N$，由假定各种类型的票券相互独立，所以得到

$$P(A_j)=\left(\frac{N-1}{N}\right)^n$$

125

如果前 n 种票券没有第 j_1 或 j_2 型则事件 $A_{j_1}A_{j_2}$ 发生．于是，利用独立性可得

$$P(A_{j_1}A_{j_2})=\left(\frac{N-2}{N}\right)^n$$

同理可得

$$P(A_{j_1}A_{j_2}\cdots A_{j_k})=\left(\frac{N-k}{N}\right)^n$$

对 $n>0$，

$$P\{T>n\}=N\left(\frac{N-1}{N}\right)^n-\binom{N}{2}\left(\frac{N-2}{N}\right)^n+\binom{N}{3}\left(\frac{N-3}{N}\right)^n-\cdots+(-1)^N\binom{N}{N-1}\left(\frac{1}{N}\right)^n$$

$$= \sum_{i=1}^{N-1} \binom{N}{i} \left(\frac{N-i}{N}\right)^n (-1)^{i+1} \tag{1-2}$$

那么，T 等于 n 的概率可以由下式得到：

$$P\{T>n-1\}=P\{T=n\}+P\{T>n\}$$

或等价地，

$$P\{T=n\}=P\{T>n-1\}-P\{T>n\}$$

现有 n 张票券，其中所含不同类型票券的个数记为随机变量 D_n. 为计算 $P\{D_n=k\}$，现固定一个有 k 种类型票券的集合，这个集合包含前 n 张票券的不同类型. 为得到这种情形，当且仅当前 n 张票券包含

A：每张是 k 种类型的一种.

B：可以获得 k 种类型的每种.

现收集到的 k 种类型票券中，某种票券的概率为 k/N，因此 A 的概率为$(k/N)^n$. 另外，给定 k 种类型中的一种票券，易见它等可能地是 k 种类型中的任意一种. 因此 A 发生的情况下，B 的条件概率与 n 张票券(其中每张票券是 k 种类型中的任意一种)中包含所有 k 种类型的票券的概率相同. 但这正是需要收集一个完整的集合的票券数的概率，当在 k 种类型中选择时，这个数目小于等于 n，于是可将式(1-2)中的 N 用 k 代替而得到. 因此有

126

$$P(A)=\left(\frac{k}{N}\right)^n$$

$$P(B \mid A) = 1-\sum_{i=1}^{k-1}\binom{k}{i}\left(\frac{k-i}{k}\right)^n(-1)^{i+1}$$

最后，有 k 种类型的集合共有$\binom{N}{k}$种，我们有

$$P\{D_n=k\}=\binom{N}{k}P(AB)=\binom{N}{k}\left(\frac{k}{N}\right)^n\left[1-\sum_{i=1}^{k-1}\binom{k}{i}\left(\frac{k-i}{k}\right)^n(-1)^{i+1}\right]$$

注 因为要获得每种类型至少有一张的整套票券，所以若 $n<N$，则 $P\{T>n\}=1$. 于是，对 $1\leqslant n<N$，由式(1-2)得到

$$\sum_{i=1}^{N-1}\binom{N}{i}\left(\frac{N-i}{N}\right)^n(-1)^{i+1}=1$$

也可以写作

$$\sum_{i=0}^{N-1}\binom{N}{i}\left(\frac{N-i}{N}\right)^n(-1)^{i+1}=0$$

在上式两边同乘$(-1)^N N^n$，并令 $j=N-i$，有

$$\sum_{j=1}^{N}\binom{N}{j}j^n(-1)^{j-1}=0 \qquad 1\leqslant n<N$$

■

对随机变量 X，函数 F 定义为

$$F(x)=P\{X\leqslant x\} \qquad -\infty<x<+\infty$$

称为**累积分布函数**(cumulative distribution function)或简称为**分布函数**. 分布函数是对于所有的实数 x，随机变量小于等于 x 的概率.

若 $a\leqslant b$，因为事件$\{X\leqslant a\}$包含在事件$\{X\leqslant b\}$中，所以前者的概率 $F(a)$小于等于后者的概

率 $F(b)$. 换言之，$F(x)$是 x 的非降函数. 分布函数的其他性质见 4.9 节.

4.2 离散型随机变量

至多能取可数多个值的随机变量称为离散型随机变量. 对于离散型随机变量 X，定义 X 的概率质量函数 $p(a)$如下：

127

$$p(a)=P\{X=a\}$$

概率质量函数 $p(a)$在至多可数个 a 上取正值. 也就是说，如果假定 X 只能取 $x_1,x_2,\cdots$中的某个值，那么有

$$\begin{cases} p(x_i)\geqslant 0 & i=1,2,\cdots \\ p(x)=0 & \text{所有其他 } x \end{cases}$$

因为 X 只取 x_i 中的某个值，故有

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

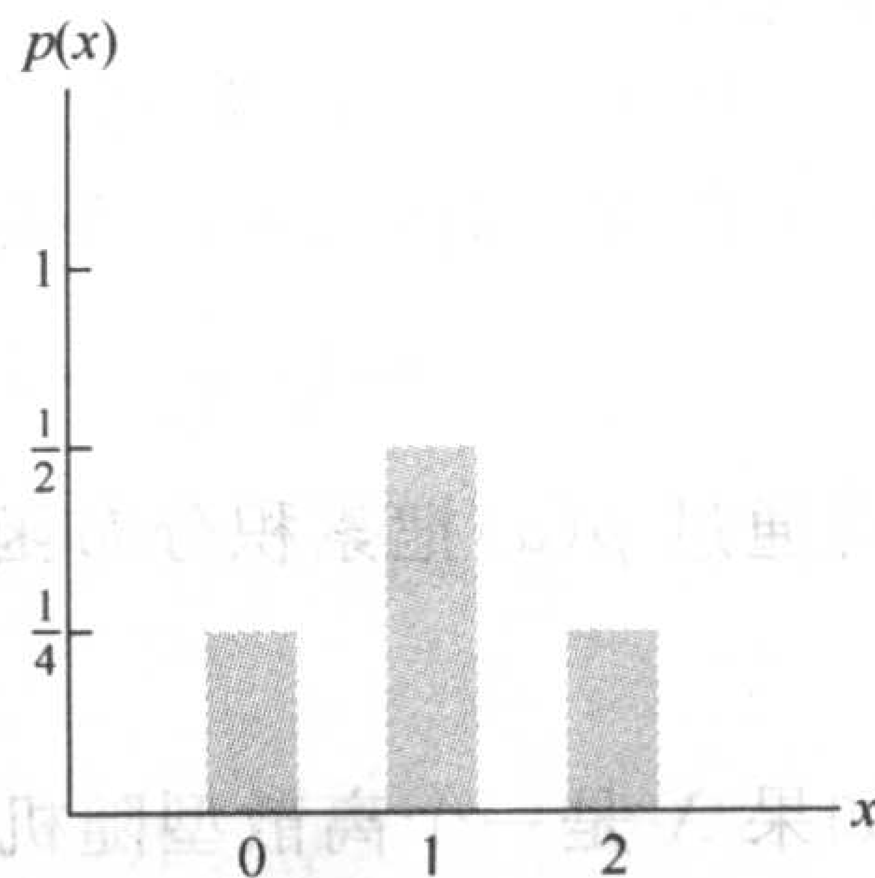


图 4-1

在 x 轴上标出 x_i，在 y 轴上标出 $p(x_i)$，可以很直观地在图上画出概率质量函数. 例如，若 X 的概率质量函数为

$$p(0)=\frac{1}{4}\quad p(1)=\frac{1}{2}\quad p(2)=\frac{1}{4}$$

则在图 4-1 中画出其概率质量函数. 同理，可以画出掷两枚骰子点数之和的概率质量函数(见图 4-2).

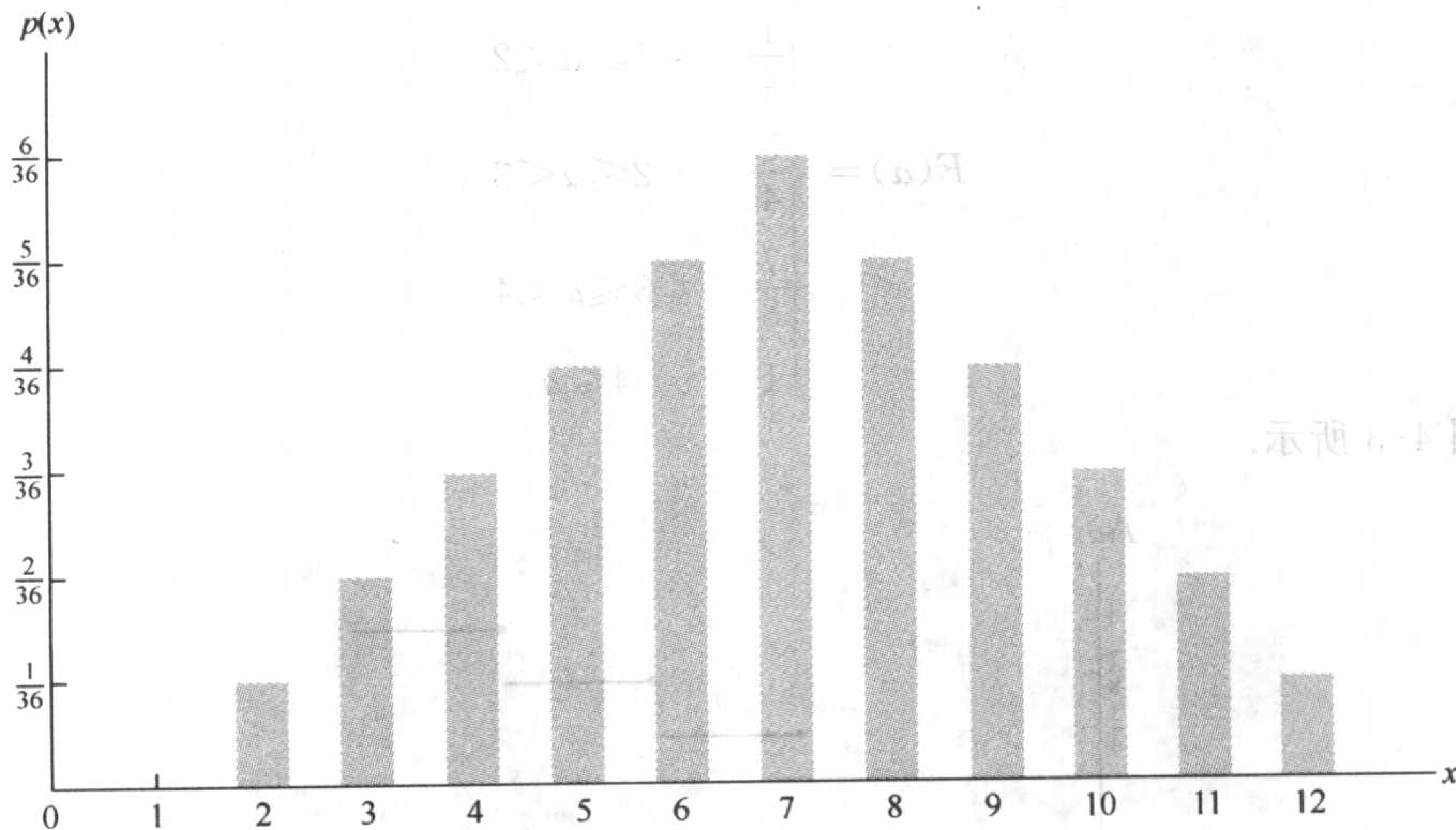


图 4-2

128

例 2a 设随机变量 X 的概率质量函数 $p(i)=c\lambda^i/i!$，其中 $i=0,1,2,\cdots,\lambda$ 为某个正值.

(a) 求 $P\{X=0\}$.

(b) 求 $P\{X>2\}$.

解 因为 $\sum_{i=0}^{\infty} p(i) = 1$，所以有

$$c\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}=1$$

利用 $e^x=\sum_{i=0}^{\infty}x^i/i!$ 可得

$$ce^{\lambda}=1 \text{ 或 } c=e^{-\lambda}$$

于是有

(a) $P\{X=0\}=e^{-\lambda}\lambda^0/0!\ =e^{-\lambda}$

(b) $P\{X>2\}=1-P\{X\leqslant 2\}=1-P\{X=0\}-P\{X=1\}-P\{X=2\}$

$$=1-e^{-\lambda}-\lambda e^{-\lambda}-\frac{1}{2}\lambda^2e^{-\lambda}$$ ■

可通过 $p(a)$ 把累积分布函数 F 表示为

$$F(a)=\sum_{\text{一切}x\leqslant a}p(x)$$

如果 X 是一个离散型随机变量，其可能值集合为 $x_1,x_2,x_3,\cdots$，其中 $x_1<x_2<x_3<\cdots$，那么 X 的分布函数 F 为阶梯函数. 这就是说，F 在区间 $[x_{i-1},x_i)$ 中取常数，且在 x_i 处有跳跃值 $p(x_i)$. 例如，若 X 的概率质量函数为

$$p(1)=\frac{1}{4},p(2)=\frac{1}{2},p(3)=\frac{1}{8},p(4)=\frac{1}{8}$$

则其累积分布函数为

$$F(a)=\begin{cases}0 & a<1\\ \frac{1}{4} & 1\leqslant a<2\\ \frac{3}{4} & 2\leqslant a<3\\ \frac{7}{8} & 3\leqslant a<4\\ 1 & 4\leqslant a\end{cases}$$

其图形如图 4-3 所示.

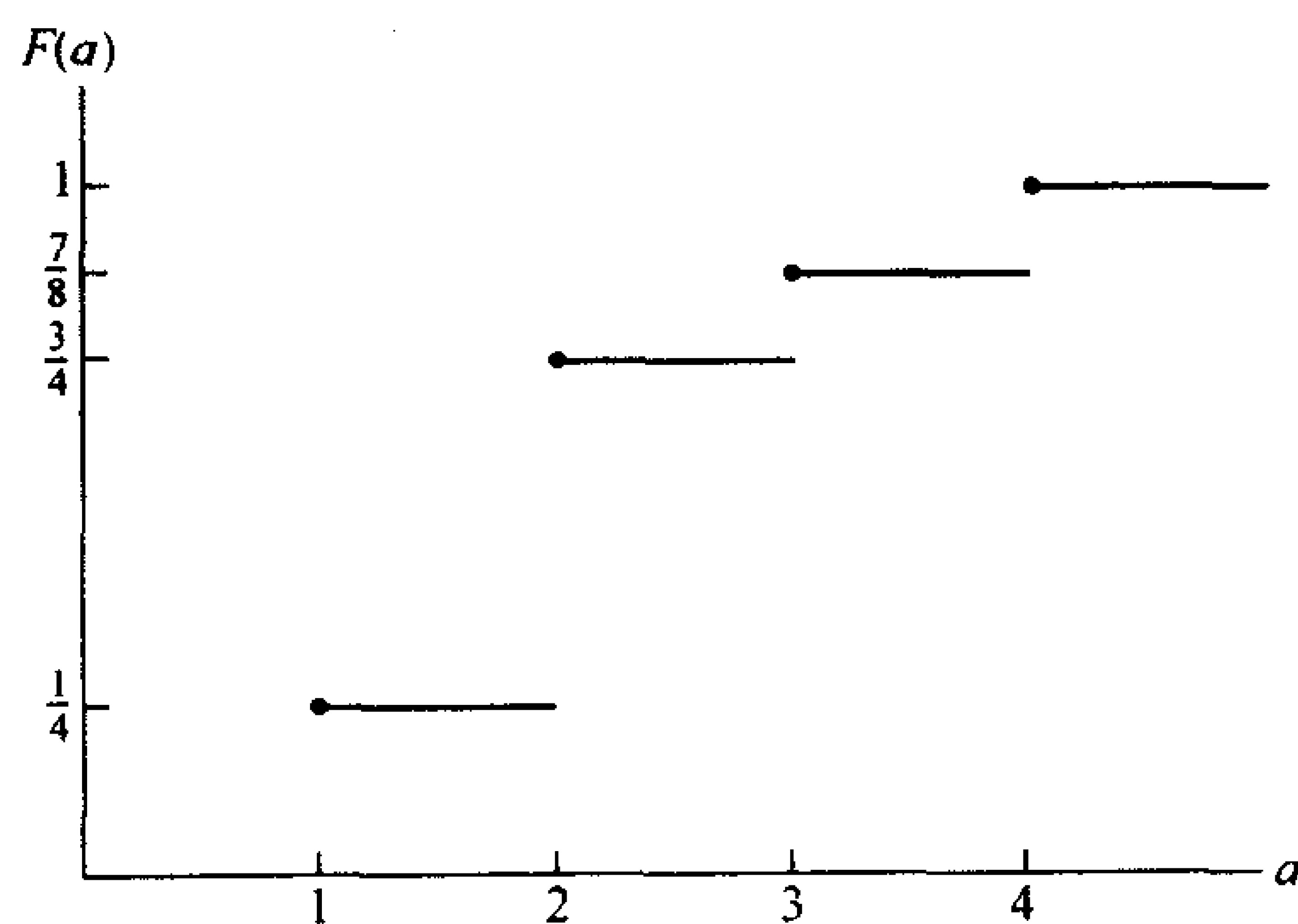


图 4-3

读者会发现，$F(a)$在1,2,3,4任意一值上的跳跃值等于X取这个值的概率.

4.3 数学期望

随机变量的数学期望是概率论中最重要的概念之一．设X为离散型随机变量，其概率质量函数为$p(x)$．定义X的数学期望或期望值$E[X]$为

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

换言之，X的期望值就是X所能取到的各个可能值的加权平均，其中每一个可能值的权重等于X取这个值的概率．例如，设X的概率质量函数为

$$p(0)=\frac{1}{2}=p(1)$$

则

$$E[X]=0\left(\frac{1}{2}\right)+1\left(\frac{1}{2}\right)=\frac{1}{2}$$

这正是X所能取的两个值0与1的普通平均．另一方面，若

$$p(0)=\frac{1}{3} \qquad p(1)=\frac{2}{3}$$

则

$$E[X]=0\left(\frac{1}{3}\right)+1\left(\frac{2}{3}\right)=\frac{2}{3}$$

这是两个可能值0与1的加权平均，由于$p(1)=2p(0)$，故加在1上的权重等于加在0上的权重的2倍.

概率的频率解释为数学期望的定义提供了另一种说明．概率的频率解释(第8章的强大数定律将给出其部分证明)是指，如果接连不断地、独立地重复某一个试验，则对任一事件E，其发生的次数所占的比例等于$P(E)$．先考虑随机变量X，它必须在$x_1,x_2,\cdots,x_n$中取值，而且取各值的相应概率为$p(x_1),p(x_2),\cdots,p(x_n)$；把$X$设想为在一局机会游戏中赢的钱数，就是说，将以概率$p(x_i)$赢得$x_i$元，$i=1,2,\cdots,n$．如果连续玩这个游戏，那么根据概率的频率解释，赢得x_i的次数所占的比例将等于$p(x_i)$．该结论对一切$i(i=1,2,\cdots,n)$都成立，因此每一局的平均赢钱数应是

129～130

$$\sum_{i=1}^{n} x_i p(x_i) = E[X]$$

例3a 设X是掷一个均匀的骰子所得的结果，试求$E[X]$.

解 因为$p(1)=p(2)=p(3)=p(4)=p(5)=p(6)=\frac{1}{6}$，所以可得

$$E[X]=1\left(\frac{1}{6}\right)+2\left(\frac{1}{6}\right)+3\left(\frac{1}{6}\right)+4\left(\frac{1}{6}\right)+5\left(\frac{1}{6}\right)+6\left(\frac{1}{6}\right)=\frac{7}{2}$$ ■

例3b 称I为事件A的示性变量，若

$$I=\begin{cases}1 & \text{若 } A \text{ 发生} \\ 0 & \text{若 } A^c \text{ 发生}\end{cases}$$

试求$E[I]$.

解 因为$p(1)=P(A),p(0)=1-P(A)$，所以有

$$E[I]=P(A)$$

也就是说，事件 A 的示性变量的期望值等于 A 发生的概率. ■

例 3c 在智力竞猜节目中，一个参与者需回答两个问题(题 1 及题 2)，他对这两个问题可以按自己选择的前后顺序去回答. 如果他决定先回答题 $i(i=1,2)$，则只有当他回答对了题 i 时才让他继续回答题 $j(j\neq i)$. 若回答错了，那么就不让他再回答另一个问题了. 如果他答对题 $i(i=1,2)$，就得奖金 V_i 美元. 因此，如果这两个问题他都答对，就得 V_1+V_2 美元. 现假定他答对题 i 的概率为 $P_i(i=1,2)$，并令 $E_i(i=1,2)$ 表示他答对题 i 的事件，这两个事件是相互独立的. 试问他应该先回答哪个问题才能使他的期望奖金最多?

解 如果他先回答题 1，则他将得

$$\begin{array}{ll} 0 & \text{以概率 } 1-P_1 \\ V_1 & \text{以概率 } P_1(1-P_2) \\ V_1+V_2 & \text{以概率 } P_1P_2 \end{array}$$

在此情形下，他的期望奖金为

$$V_1P_1(1-P_2)+(V_1+V_2)P_1P_2$$

另一方面，如果他先回答题 2，则他的期望奖金为

131

$$V_2P_2(1-P_1)+(V_1+V_2)P_1P_2$$

因此，如果

$$V_1P_1(1-P_2)\geqslant V_2P_2(1-P_1)$$

或等价地，如果

$$\frac{V_1P_1}{1-P_1}\geqslant\frac{V_2P_2}{1-P_2}$$

那么他最好先回答题 1. 因此，例如，若他有 60% 的把握答对题 1，题 1 的奖金为 200 元，而有 80% 的把握答对题 2，题 2 的奖金为 100 元，则他应该先回答题 2，这是因为

$$400=\frac{(100)(0.8)}{0.2}>\frac{(200)(0.6)}{0.4}=300$$ ■

例 3d 120 名学生乘坐 3 辆公交车去听音乐会，其中 36、40、44 名学生各乘一辆车. 120 名学生随机上车，设 X 表示每辆车上的学生人数，试求 $E[X]$.

解 由于是从 120 名学生中等可能地随机选择学生，所以有

$$P\{X=36\}=\frac{36}{120}\qquad P\{X=40\}=\frac{40}{120}\qquad P\{X=44\}=\frac{44}{120}$$

于是

$$E[X]=36\left(\frac{3}{10}\right)+40\left(\frac{1}{3}\right)+44\left(\frac{11}{30}\right)=\frac{1\,208}{30}=40.266\,7$$

另一方面，每辆车上的平均学生人数为 $120/3=40$，说明随机地选择学生的公交车上的学生人数的期望值大于一辆公交车上学生人数的平均值. 这是一个普遍的现象，出现这种现象的原因是一辆公交车上可以乘坐的学生人数越多，学生就越有可能乘坐这辆公交车. 结果是学生人数多的汽车比学生人数少的汽车给出的权重大(参见自测题 4). ■

注 数学期望的概念类似于物理学中质量分布的重心(center of gravity). 考虑离散型随机变量 X，其概率质量函数为 $p(x_i),i\geqslant 1$. 若我们现在设想一根重量很轻的杆置于 $x_i(i\geqslant 1)$，其质量为 $p(x_i)(i\geqslant 1)$ (见图 4-4). 杆的平衡点称为重心. 熟悉初等数理统

计的读者很容易证明该点就是 $E[X]$[⊖].

[132]

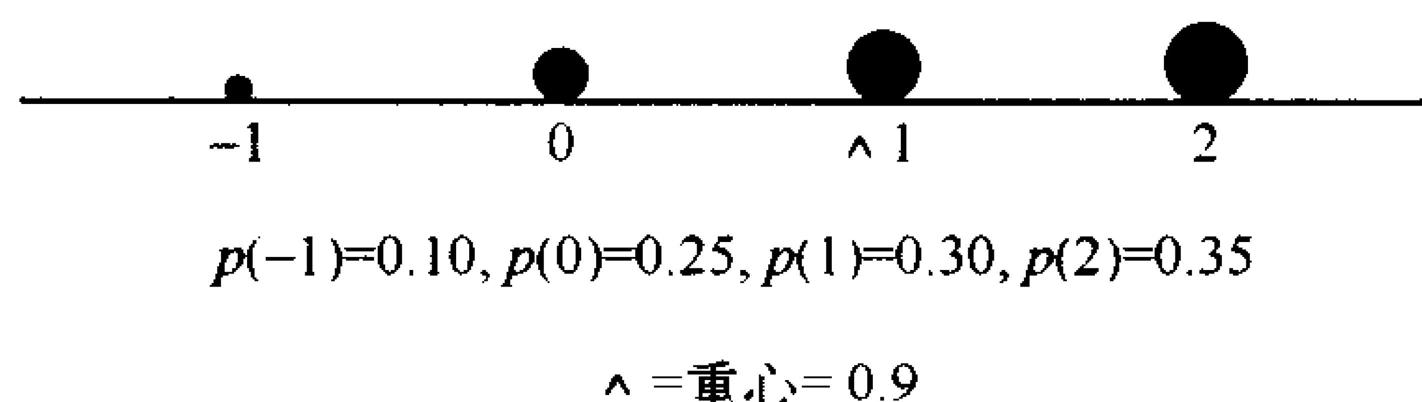


图 4-4

4.4 随机变量函数的数学期望

设已知离散型随机变量 X 及其概率质量函数，我们需要计算的是 X 的某一函数的数学期望，比如说 $g(X)$的数学期望. 应该怎样来做呢？有一种方法如下：因为 $g(X)$也是离散型随机变量，故应有概率质量函数，这可由 X 的概率质量函数求出. 一旦知道 $g(X)$的概率质量函数，就可以根据数学期望的定义把 $E[g(X)]$计算出来.

例 4a 设 X 是取值为 -1，0，1 的随机变量，其概率质量函数为

$$P\{X=-1\}=0.2 \qquad P\{X=0\}=0.5 \qquad P\{X=1\}=0.3$$

试求 $E[X^2]$.

解 令 $Y=X^2$，则 Y 的概率质量函数为

$$P\{Y=1\}=P\{X=-1\}+P\{X=1\}=0.5$$
$$P\{Y=0\}=P\{X=0\}=0.5$$

所以

$$E[X^2]=E[Y]=1(0.5)+0(0.5)=0.5$$

读者需注意

$$0.5=E[X^2]\neq(E[X])^2=0.01$$ ■

尽管用上述方法总能从已知 X 的概率质量函数算出 X 的任一函数的数学期望，但还有一种更容易的方法计算 $E[g(X)]$. 也就是说，当 X 等于 x 时，$g(X)$等于 $g(x)$. $E[g(X)]$可看成是 $g(x)$的加权平均，$g(x)$的权重是 X 等于 x 时 $g(x)$的概率. 因此，下面的结论是很直观的.

[133]

命题 4.1 *若 X 为离散型随机变量，以概率质量函数 $p(x_i)$取值为 $x_i(i\geqslant 1)$，则对任一实值函数 g，有*

$$E[g(X)]=\sum_i g(x_i)p(x_i)$$

在证明这个命题之前，先验证一下该命题与例 4a 的结果是一致的. 将此命题应用到例 4a 得

$$E\{X^2\}=(-1)^2(0.2)+0^2(0.5)+1^2(0.3)=1(0.2+0.3)+0(0.5)=0.5$$

与前面所得的结果一致.

命题 4.1 的证明 为了证明命题 4.1，正如前面验证的方法一样，将 $\sum_i g(x_i)p(x_i)$ 中 $g(x_i)$ 的值相同的所有项分为一组. 特别假定 $y_j(j\geqslant 1)$表示 $g(x_i)(i\geqslant 1)$的不同值，则有相同值的全部 $g(x_i)$的组给出

⊖ 为了证明这一点，我们必须证明围绕 $E[X]$的点的力矩的和等于零，即需要证明 $0=\sum_i(x_i-E[X])p(x_i)$，这很容易得到.

$$\sum_i g(x_i)p(x_i) = \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) = \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) = \sum_j y_j P\{g(X)=y_j\} = E[g(X)] \quad \blacksquare$$

例 4b 按季节出售的某种应时商品，每出售一件获得纯利润 b 美元，如果到季末还有剩余商品未能售完，则每件将净亏损ℓ美元. 设在任一季节内，在某特定的百货商店该商品被订购的件数是一个随机变量，其概率质量函数为 $p(i), i\geqslant 0$. 如果这个商店必须提前储备该商品，试求为使商店获得最大的期望利润，应储备多少件该商品?

解 设 X 表示订购该商品的件数. 如果储备 s 件，其利润记为 $P(s)$，则 $P(s)$可表示为

$$P(s) = \begin{cases} bX-(s-X)\ell & X\leqslant s \\ sb & X>s \end{cases}$$

134

故其期望利润为

$$\begin{aligned} E[P(s)] &= \sum_{i=0}^{s}[bi-(s-i)\ell]p(i) + \sum_{i=s+1}^{\infty} sbp(i) \\ &= (b+\ell)\sum_{i=0}^{s} ip(i) - s\ell\sum_{i=0}^{s}p(i) + sb\left[1-\sum_{i=0}^{s}p(i)\right] \\ &= (b+\ell)\sum_{i=0}^{s} ip(i) - (b+\ell)s\sum_{i=0}^{s}p(i) + sb \\ &= sb + (b+\ell)\sum_{i=0}^{s}(i-s)p(i) \end{aligned}$$

为确定 s 的最优值，先研究当 s 增加 1 件时所获得的利润有什么变化. 经变量替换，此时的期望利润等于

$$\begin{aligned} E[P(s+1)] &= b(s+1) + (b+\ell)\sum_{i=0}^{s+1}(i-s-1)p(i) \\ &= b(s+1) + (b+\ell)\sum_{i=0}^{s}(i-s-1)p(i) \end{aligned}$$

所以

$$E[P(s+1)] - E[P(s)] = b - (b+\ell)\sum_{i=0}^{s}p(i)$$

因此，只要

$$\sum_{i=0}^{s}p(i) < \frac{b}{b+\ell} \tag{4-1}$$

则储备 $s+1$ 件将比储备 s 件更好. 因为式(4-1)的左边是 s 的递增函数而右边是常数，故不等式(4-1)对一切 $s\leqslant s^*$ 成立，其中 s^* 为满足式(5-1)的最大 s 值. 由于

$$E[P(0)]<\cdots<E[P(s^*)]<E[P(s^*+1)]>E[P(s^*+2)]>\cdots$$

因而，储备 s^*+1 件商品可使期望利润达到最大值. $\blacksquare$

例 4c(有效性) 假设你必须选择两种可能行动中的一种，每种行动都会产生 n 个结果中的任意一个，这 n 个结果分别记为 $C_1,\cdots,C_n$. 如果选择第 1 种行动，则以概率 $p_i(i=1,\cdots,n)$ 产生结果 C_i；选择第 2 种行动，则以概率 $q_i(i=1,\cdots,n)$产生结果 C_i，其中 $\sum_{i=1}^{n}p_i = \sum_{i=1}^{n}q_i = 1$. 用下面的方法来确定该选择哪种行动. 先给不同结果赋数值. 首先，确定最想要的结果和最不想要的结果，分别记为 C 和 c. 给结果 c 赋值 0，给结果 C 赋值 1. 现在考虑其余 $n-2$ 种结果，比如说 C_i. 为给 C_i 赋值，设想你或者选择能得到结果 C_i 的那种行动，或者做一个随机试验，该试验或以概率 u 得到结果 C，或以概率 $1-u$ 得到结果 c. 显然，你的选择取决于 u 的值. 若

135

$u=1$，试验一定有结果 C. 因为 C 是最想要的结果，你于是会选择结果为 C_i 的试验. 另一方面，若 $u=0$，试验一定得到最不想要的结果，即 c，这种情形下你会选择结果为 C_i 的试验. 现在，若 u 从 1 降到 0，在某一个点上，你选择将试验转换到某一轮 C_i 似乎是合理的，且在该临界点这两种选择是无差别的. 取无差别的概率 u 作为结果 C_i 的值. 换句话说，取结果 C_i 的值为 u，你或者选择能得到结果 C_i 的行动，或者做一个试验，以概率 u 得到结果 C，或以概率 $1-u$ 得到结果 c. 结果 C_i 的有效性(utility)记为 $u(C_i)$，我们称之为无差别概率.

为确定哪种行动是最优的，可以对每种行动求值. 先考虑第 1 种行动，以概率 p_i 出现结果 $C_i, i=1,\cdots,n$. 可以通过两步得到结果. 第一步，根据概率 $p_1,\cdots,p_n$ 赋值 $1,\cdots,n$ 中的一个；若赋值 i，则会得到结果 C_i. 但是，由于结果 C_i 等价于以概率 $u(C_i)$ 得到结果 C 或以概率 $1-u(C_i)$ 得到结果 c，因此两步试验的结果等价于进行一个试验，可能得到结果 C 或 c，以概率

$$\sum_{i=1}^{n} p_i u(C_i)$$

得到结果 C.

同理，选择第 2 种行动等价于参加一个试验，可能得到结果 C 或 c，以概率

$$\sum_{i=1}^{n} q_i u(C_i)$$

得到结果 C.

因为 C 比 c 更想得到，所以若

$$\sum_{i=1}^{n} p_i u(C_i) > \sum_{i=1}^{n} q_i u(C_i)$$

则第 1 种行动比第 2 种行动好. 换言之，行动可以通过结果的有效性的期望值来衡量，有最大可能期望值的行动是较好的选择. ■

136

命题 4.1 的一个简单推论为推论 4.1.

推论 4.1 *若 a 和 b 是常数，则*

$$E[aX+b]=aE[X]+b$$

证明

$$E[aX+b] = \sum_{x:p(x)>0} (ax+b)p(x) = a\sum_{x:p(x)>0} xp(x) + b\sum_{x:p(x)>0} p(x) = aE[X]+b \quad ■$$

随机变量 X 的数学期望 $E[X]$ 也称为 X 的均值或一阶矩. 量 $E[X^n]$($n\geqslant 1$)称为 X 的 n 阶矩. 由命题 4.1 可知

$$E[X^n] = \sum_{x:p(x)>0} x^n p(x)$$

4.5 方差

给定一个随机变量 X，其分布函数为 F，用合适的工具来描述 F 的基本性质是极为有用的. 可以通过 $E[X]$，即 X 的期望来描述. 然而 $E[X]$ 是 X 的可能值的加权平均，它不能说明这些值的分散程度. 例如，随机变量 W，Y 与 Z 的概率质量函数定义如下：

$$W=0 \quad 以概率 1$$

$$Y=\begin{cases} -1 & 以概率\dfrac{1}{2} \\ +1 & 以概率\dfrac{1}{2} \end{cases}$$

$$Z=\begin{cases}-100 & \text{以概率}\frac{1}{2}\\ +100 & \text{以概率}\frac{1}{2}\end{cases}$$

137

随机变量 W，Y 与 Z 有相同的期望，均为 0，而 Y 分散得比 W 大，Z 分散得比 Y 大.

因为我们期望 X 在均值 $E[X]$ 周围取值，因此就出现了如何合理地度量 X 的可能变差的问题，即 X 与其均值的偏离程度. 度量变差的一种方法是考虑量 $E[|X-\mu|]$ 的大小，这里 $\mu=E[X]$. 但是，在数学上处理这个量不方便，因此，通常考虑一个更易处理的量，即 X 与它的均值之差的平方的数学期望. 从而，我们有如下定义.

定义 假定随机变量 X 的均值为 μ，那么 X 的方差(记为 $\mathrm{Var}(X)$)定义为

$$\mathrm{Var}(X)=E[(X-\mu)^2]$$

方差 $\mathrm{Var}(X)$ 的另一个计算公式推导如下：

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x) = \sum_x (x^2-2\mu x+\mu^2)p(x)\\ &= \sum_x x^2 p(x) - 2\mu\sum_x xp(x) + \mu^2\sum_x p(x)\\ &= E[X^2]-2\mu^2+\mu^2 = E[X^2]-\mu^2\end{aligned}$$

即

$$\boxed{\mathrm{Var}(X)=E[X^2]-(E[X])^2}$$

简言之，X 的方差等于 X^2 的期望值减去 X 的期望值的平方. 实际上，这也是计算 $\mathrm{Var}(X)$ 最简单的方法.

例 5a 设 X 是掷一个均匀的骰子所得到的点数，试求 $\mathrm{Var}(X)$.

解 在例 3a 中，已知 $E[X]=\frac{7}{2}$，且

138

$$E[X^2]=1^2\left(\frac{1}{6}\right)+2^2\left(\frac{1}{6}\right)+3^2\left(\frac{1}{6}\right)+4^2\left(\frac{1}{6}\right)+5^2\left(\frac{1}{6}\right)+6^2\left(\frac{1}{6}\right)=\left(\frac{1}{6}\right)(91)$$

因此

$$\mathrm{Var}(X)=\frac{91}{6}-\left(\frac{7}{2}\right)^2=\frac{35}{12}$$ ■

对常数 a 与 b 恒有

$$\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X)$$

为证明上式，令 $\mu=E[X]$，由推论 4.1 可知 $E[aX+b]=a\mu+b$，所以

$$\mathrm{Var}(aX+b)=E[(aX+b-a\mu-b)^2]=E[a^2(X-\mu)^2]=a^2E[(X-\mu)^2]=a^2\mathrm{Var}(X)$$

注 (a) 可把均值看成是质量分布的重心，而方差则表示力学中的转动惯量.

(b) $\mathrm{Var}(X)$ 的平方根称为 X 的标准差，记为 $\mathrm{SD}(X)$. 即

$$\mathrm{SD}(X)=\sqrt{\mathrm{Var}(X)}$$

离散型随机变量总是根据其概率质量函数来归类，下面几节讨论一些常用的随机变量.

4.6 伯努利随机变量与二项随机变量

进行一次试验，其结果要么是“成功”，要么是“失败”. 如果结果是成功时令 $X=1$，而结果是失败时令 $X=0$，那么 X 的概率质量函数为

$$p(0)=P\{X=0\}=1-p$$
$$p(1)=P\{X=1\}=p \tag{6-1}$$

其中 $p(0\leqslant p\leqslant 1)$是试验结果为“成功”的概率.

随机变量 X 称为伯努利随机变量(来自瑞士数学家詹姆士・伯努利)，如果其概率质量函数由式(6-1)给出，其中 $p\in(0,1)$.

现进行 n 次独立试验，每次试验“成功”的概率为 p，而“失败”的概率为 $1-p$. 若以 X 表示 n 次试验中成功的次数，则 X 称为以(n,p)为参数的二项随机变量(binomial random variable). 因此，伯努利随机变量是以$(1,p)$为参数的二项随机变量.

139

以(n,p)为参数的二项随机变量的概率质量函数为

$$p(i)=\binom{n}{i}p^i(1-p)^{n-i} \qquad i=0,1,2,\cdots,n \tag{6-2}$$

式(6-2)的正确性可以这样验证：首先，由试验的独立性假设可知，由 i 次成功、$n-i$ 次失败这 n 个结果组成的任一特定序列出现的概率为 $p^i(1-p)^{n-i}$. 再者，这种不同的序列共有 $\binom{n}{i}$ 个，于是得到式(6-2). 这也许最容易从 n 次实验中 i 次成功共有 $\binom{n}{i}$ 种方式看出. 例如，若 $n=4$，$i=2$ 则四次试验中两次成功共有 $\binom{4}{2}=6$ 种方式，即(s,s,f,f)，(s,f,s,f)，(s,f,f,s)，(f,s,s,f)，(f,s,f,s)，(f,f,s,s)，这里，例如(s,s,f,f)表示前两次试验成功而后两次试验失败. 因为每个这样的结果发生的概率均为 $p^2(1-p)^2$，于是所求四次试验中有两次试验成功的概率为 $\binom{4}{2}p^2(1-p)^2$.

运用二项式定理，式(6-2)给出的概率对 i 求和等于 1，即

$$\sum_{i=0}^{\infty}p(i)=\sum_{i=0}^{n}\binom{n}{i}p^i(1-p)^{n-i}=[p+(1-p)]^n=1$$

例 6a 抛 5 枚均匀的硬币，如果假定各硬币抛的结果相互独立，试求所得正面个数的概率质量函数.

解 令 X 表示正面(成功)出现的个数，那么 X 是以$\left(n=5,p=\frac{1}{2}\right)$为参数的二项随机变量，这样由式(6-2)可得

$$P\{X=0\}=\binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5=\frac{1}{32}$$

$$P\{X=1\}=\binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4=\frac{5}{32}$$

$$P\{X=2\}=\binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3=\frac{10}{32}$$

$$P\{X=3\}=\binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2=\frac{10}{32}$$

$$P\{X=4\}=\binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1=\frac{5}{32}$$

140

$$P\{X=5\}=\binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0=\frac{1}{32}$$ ■

例 6b 已知某公司生产的螺丝以 0.01 的概率为次品，并设各个螺丝是否为次品是相互独立的. 这家公司将每 10 个螺丝包成一包出售，并保证若发现某包内多于一个次品，则可退款. 问卖出的各包螺丝中，被退回公司的占多大比例？

解 令 X 表示某包螺丝中次品的个数，则 X 为以(10,0.01)为参数的二项随机变量. 因此，这包螺丝被退回的概率为

$$1-P\{X=0\}-P\{X=1\}=1-\binom{10}{0}(0.01)^0(0.99)^{10}-\binom{10}{1}(0.01)^1(0.99)^9\approx 0.004$$

于是，将被退回的包数仅占 0.4%. ■

例 6c 下述赌博游戏称为“幸运轮”(或掷骰赌博)，在狂欢节和许多赌博俱乐部里十分流行. 赌徒将赌注押在 1 至 6 的某一数上，然后掷三个骰子，约定若赌徒所押的数出现 i 次，$i=1,2,3$，则赌徒赢 i 元；另一方面，如果赌徒所押的数没有出现，则他输 1 元. 问这个规则对赌徒是否公平？(实际上，幸运轮是旋转一个轮盘，在标记了数字 1～6 的 3 个数字处停止移动，在数学上它等价于掷骰赌博.)

解 如果假定这些骰子是均匀的，且掷出的点数相互独立，则赌徒所押的数在骰子上出现的次数是以$\left(3,\frac{1}{6}\right)$为参数的二项随机变量. 这样，若以 X 表示赌徒赢的钱数，则有

$$P\{X=-1\}=\binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3=\frac{125}{216}$$

$$P\{X=1\}=\binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2=\frac{75}{216}$$

$$P\{X=2\}=\binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1=\frac{15}{216}$$

$$P\{X=3\}=\binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0=\frac{1}{216}$$

为确定对赌徒而言这是不是公平赌博，可根据前面的概率得到 $E[X]$

$$E[X]=\frac{-125+75+30+3}{216}=\frac{-17}{216}$$

141

也就是说，他每赌 216 局要平均输掉 17 元，从这个观点看，显然赌博是不公平的. ■

在下一个例子中，将考虑孟德尔(G. Mendel，1822—1884，奥地利遗传学家)所发现的遗传理论的最简单形式.

例 6d 设人的某指定特征(如眼睛的颜色、习惯用左手，等等)是由他的一对基因所决定的. 以 d 表示显性基因，r 表示隐性基因，则具有 dd 基因的人是纯显性的，具有 rr 基因的人是纯隐性的，而具有 rd 基因的人是混合性的. 纯显性与混合性的人都表现出显性基因决定的特征. 孩子从他父母身上各得到 1 个基因，现在针对某一指定特征而言，如果夫妻二人都是混合性的，他们有 4 个孩子，那么其中 3 个孩子具有显性基因所决定的特征的概率是多少？

解 假定每个孩子从他父母的两个基因中继承哪一个是等可能的. 在父母都是混合性的情况下，孩子的一对基因是 dd，rr，rd 的概率分别为$\frac{1}{4}$，$\frac{1}{4}$，$\frac{1}{2}$. 由于当孩子的一对基因为 dd 或

rd 时他具有显性基因所决定的特征，因此有这一特征的孩子的个数是参数为$\left(4,\frac{3}{4}\right)$的二项随机变量. 故所求的概率为

$$\binom{4}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1=\frac{27}{64}$$ ■

例 6e 某陪审团由 12 名陪审员组成，至少要有 8 名陪审员投有罪的票才能宣判被告有罪. 假设陪审员的判断是相互独立的，且每位陪审员做出正确判断的概率为 θ，问这个陪审团做出正确判决的概率是多少?

解 所述的问题是无法求解的，因为还缺少足够的条件. 例如，若被告是无罪的，则陪审团做出正确判决的概率为

$$\sum_{i=5}^{12}\binom{12}{i}\theta^i(1-\theta)^{12-i}$$

而若被告有罪，正确判决的概率为

$$\sum_{i=8}^{12}\binom{12}{i}\theta^i(1-\theta)^{12-i}$$

因此若以 α 表示被告有罪的概率，那么以他是不是有罪为条件，得到陪审团做出正确判决的概率为

$$\alpha\sum_{i=8}^{12}\binom{12}{i}\theta^i(1-\theta)^{12-i}+(1-\alpha)\sum_{i=5}^{12}\binom{12}{i}\theta^i(1-\theta)^{12-i}$$ ■ 142

例 6f 某通信设备由 n 个元件组成，每个元件是相互独立的，而且每个元件运行的概率为 p，如果至少一半的元件运行则整个设备能够有效工作.

(a) 试问 p 取什么值时，5 个元件的设备比 3 个元件的设备更可能有效工作?

(b) 一般地，p 取什么值时，有$(2k+1)$个元件的设备比有$(2k-1)$个元件的设备更容易有效工作?

解 (a)正常运行的元件数是以参数为(n,p)的二项随机变量，5 个元件的设备有效工作的概率为

$$\binom{5}{3}p^3(1-p)^2+\binom{5}{4}p^4(1-p)+p^5$$

而 3 个元件的设备有效工作的对应概率为

$$\binom{3}{2}p^2(1-p)+p^3$$

因此，若

$$10p^3(1-p)^2+5p^4(1-p)+p^5>3p^2(1-p)+p^3$$

说明 5 个元件的设备好.

化简得

$$3(p-1)^2(2p-1)>0$$

或

$$p>\frac{1}{2}$$

(b) 一般地，$(2k+1)$个元件的设备比$(2k-1)$个元件的设备更容易有效工作当(且仅当)

$p>\frac{1}{2}$. 为证明这个结论，先考虑$(2k+1)$个元件的设备，设 X 表示前$(2k-1)$个元件正常工作的元件数，则

$$P_{2k+1}(\text{有效})=P\{X\geqslant k+1\}+P\{X=k\}(1-(1-p)^2)+P\{X=k-1\}p^2$$

$(2k+1)$个元件的设备能有效工作，如果

(i) $X\geqslant k+1$.

(ii) $X=k$，而且其余 2 个元件至少 1 个元件正常工作.

(iii) $X=k-1$，并且其余 2 个元件正常工作.

由于

143

$$P_{2k-1}(\text{有效})=P\{X\geqslant k\}=P\{X=k\}+P\{X\geqslant k+1\}$$

可以得到

$$\begin{aligned}P_{2k+1}(\text{有效})-P_{2k-1}(\text{有效})&=P\{X=k-1\}p^2-(1-p)^2P\{X=k\}\\&=\binom{2k-1}{k-1}p^{k-1}(1-p)^kp^2-(1-p)^2\binom{2k-1}{k}p^k(1-p)^{k-1}\\&=\binom{2k-1}{k}p^k(1-p)^k[p-(1-p)]\qquad\text{因为}\binom{2k-1}{k-1}=\binom{2k-1}{k}\\&>0\Leftrightarrow p>\frac{1}{2}\end{aligned}$$

■

4.6.1 二项随机变量的性质

下面介绍参数为(n,p)的二项随机变量的性质，从计算数学期望和方差开始.

$$E[X^k]=\sum_{i=0}^{n}i^k\binom{n}{i}p^i(1-p)^{n-i}=\sum_{i=1}^{n}i^k\binom{n}{i}p^i(1-p)^{n-i}$$

由恒等式

$$i\binom{n}{i}=n\binom{n-1}{i-1}$$

得到

$$\begin{aligned}E[X^k]&=np\sum_{i=1}^{n}i^{k-1}\binom{n-1}{i-1}p^{i-1}(1-p)^{n-i}\\&=np\sum_{j=0}^{n-1}(j+1)^{k-1}\binom{n-1}{j}p^j(1-p)^{n-1-j}\qquad\text{令 } j=i-1\\&=npE[(Y+1)^{k-1}]\end{aligned}$$

其中 Y 是参数为$(n-1,p)$的二项随机变量. $k=1$ 时上式变为

$$E[X]=np$$

这就是说，若每次试验成功的概率为 p，则 n 次独立试验成功的次数就等于 np. $k=2$ 时上面的公式变为

$$E[X^2]=npE[Y+1]=np[(n-1)p+1]$$

144

因为 $E[X]=np$，我们可以得到

$$\text{Var}(X)=E[X^2]-(E[X])^2=np[(n-1)p+1]-(np)^2=np(1-p)$$

综上所述，有如下结论：

若 X 是参数为 (n,p) 的二项随机变量，则

$$E[X]=np$$
$$\mathrm{Var}(X)=np(1-p)$$

下面的命题说明二项随机变量的概率质量函数先递增后递减.

命题 6.1 如果 X 是参数为 (n,p) 的二项随机变量，其中 $0<p<1$，那么，当 k 由 0 变到 n 时，$P\{X=k\}$ 先是单调增加、随后单调减少，而且当 k 为不大于 $(n+1)p$ 的最大整数时，$P\{X=k\}$ 达到最大值.

证明 为证命题，考虑怎样的 k 值使

$$P\{X=k\}/P\{X=k-1\}$$

大于或小于 1. 因为

$$\frac{P\{X=k\}}{P\{X=k-1\}}=\frac{\dfrac{n!}{(n-k)!\ k!}p^{k}(1-p)^{n-k}}{\dfrac{n!}{(n-k+1)!\ (k-1)!}p^{k-1}(1-p)^{n-k+1}}=\frac{(n-k+1)p}{k(1-p)}$$

所以，$P\{X=k\}\geqslant P\{X=k-1\}$ 当且仅当

$$(n-k+1)p\geqslant k(1-p)$$

或等价地，

$$k\leqslant(n+1)p$$

命题得证. ■

145

作为命题 6.1 的一个直观解释，图 4-5 画出了以 $\left(10,\frac{1}{2}\right)$ 为参数的二项随机变量的概率质量函数图形.

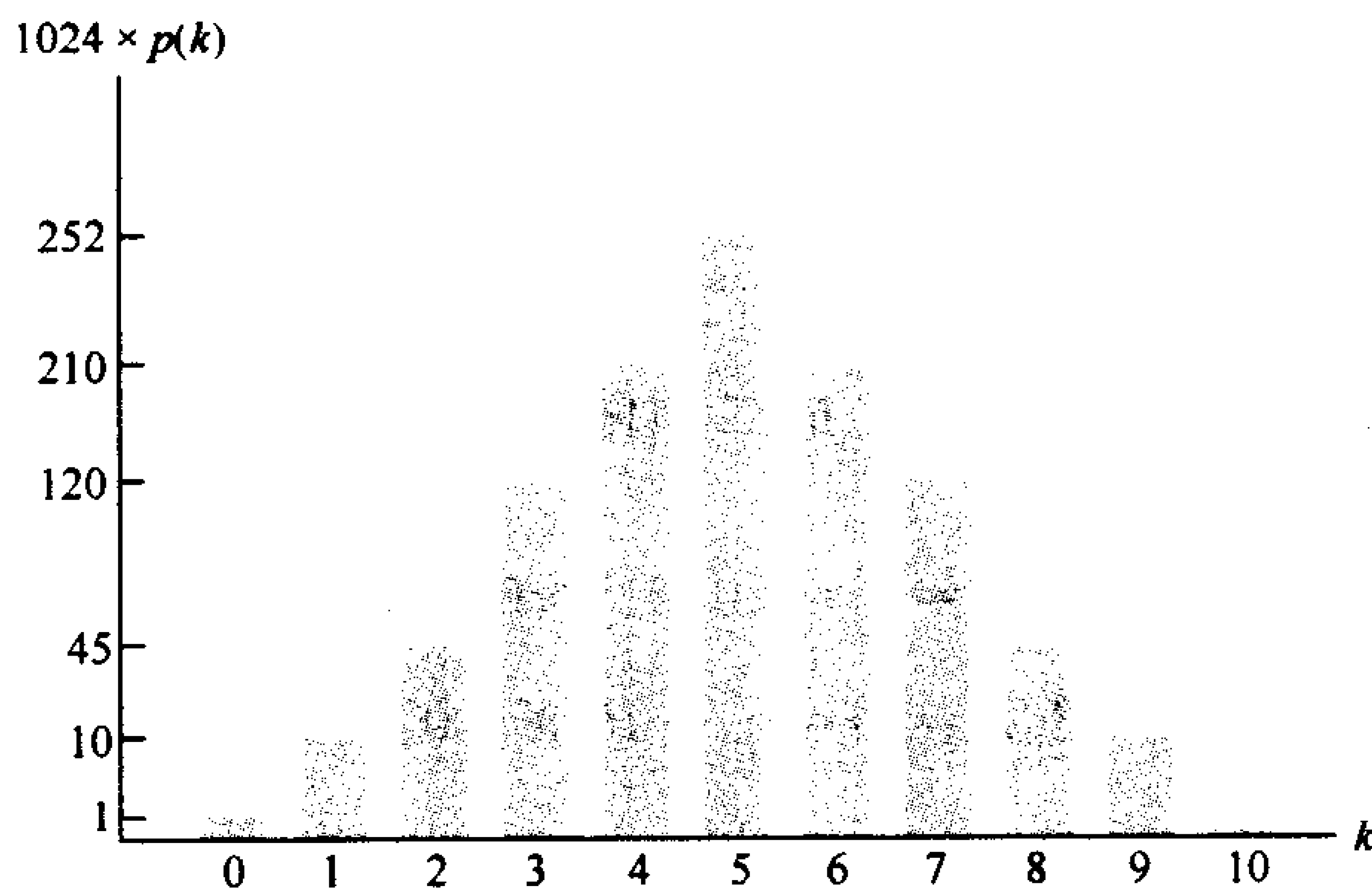


图 4-5 $p(k)=\binom{10}{k}\left(\frac{1}{2}\right)^{10}$ 的图形

例 6g 在美国总统选举中，在某个州赢得最多选票的候选人将会得到分配给该州的选举团选票总数的奖励. 对某个固定的州，选举团选票的数量与该州的人口数量粗略地成正比，即如果某个州有 n 个人，则粗略地有 nc 张选票.（实际上，它接近于 $nc+2$，因为这个州的众议院的每个成员一张选票，而参加投票的人数粗略地与该州的人口数量成正比，另外两个参议员每人一张选票.）我们计算某个人口数为 n 的州的总统选举的平均权力，其中平均权力(average

power)可以如下表示：某州有人口 $n=2k+1$，如果 $n-1$ 个人投两个候选人的选票一样，那么有一个人的票将会起决定作用.（这里我们考虑 n 为奇数，n 为偶数的情形类似.）选举结束了，我们假定 $n-1=2k$ 张选票的每张相互独立且等可能地投给每个候选人. 因此人口为 $n=2k+1$ 的某州，有一个人的票对选举结果的概率有影响，这和抛 $2k$ 枚均匀硬币出现正面和反面的次数相同的概率计算是一样的. 即

$$P\{\text{某有 } 2k+1 \text{ 个选民的州做出选择}\}=\binom{2k}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^k=\frac{(2k)!}{k!\ k!\ 2^{2k}}$$

146

运用斯特林逼近公式，当 k 很大时，

$$k!\ \sim k^{(k+1)/2}\mathrm{e}^{-k}\sqrt{2\pi}$$

记号 $a_k\sim b_k$ 表示 $\lim_{k\to\infty}a_k/b_k=1$. 因此

$$P\{\text{某有 } 2k+1 \text{ 个选民的州做出选择}\}\sim\frac{(2k)^{(2k+1)/2}\mathrm{e}^{-2k}\sqrt{2\pi}}{k^{2k+1}\mathrm{e}^{-2k}(2\pi)2^{2k}}=\frac{1}{\sqrt{k\pi}}$$

假如他做不同的选择，则这个选民的选票将影响 nc 张选票，我们看到选民数为 n 的州的选票数的期望值将受到影响，即选民的平均权力受到影响，选民的平均权力如下：

$$\text{平均权力}=ncP\{\text{做出选择}\}\sim\frac{nc}{\sqrt{n\pi/2}}=c\sqrt{2n/\pi}$$

因此，某州人口数为 n，平均权力正比于 n 的平方根. 这说明在总统选举中，人口多的州有比较大的平均权力. ■

4.6.2 计算二项分布函数

设 X 是以参数为 (n,p) 的二项随机变量，计算其分布函数

$$P\{X\leqslant i\}=\sum_{k=0}^{i}\binom{n}{k}p^k(1-p)^{n-k}\qquad i=0,1,\cdots,n$$

的关键是利用 $P\{X=k+1\}$ 与 $P\{X=k\}$ 之间的关系，这一关系由命题 6.1 的证明中得到：

$$P\{X=k+1\}=\frac{p}{1-p}\frac{n-k}{k+1}P\{X=k\}\tag{6-3}$$

例 6h 令 X 是参数为 $n=6$，$p=0.4$ 的二项随机变量，因为 $P\{X=0\}=(0.6)^6$，运用式(6-3)递推得到

$$P\{X=0\}=(0.6)^6\approx0.046\,7$$

147

$$P\{X=1\}=\frac{4}{6}\frac{6}{1}P\{X=0\}\approx0.186\,6$$

$$P\{X=2\}=\frac{4}{6}\frac{5}{2}P\{X=1\}\approx0.311\,0$$

$$P\{X=3\}=\frac{4}{6}\frac{4}{3}P\{X=2\}\approx0.276\,5$$

$$P\{X=4\}=\frac{4}{6}\frac{3}{4}P\{X=3\}\approx0.138\,2$$

$$P\{X=5\}=\frac{4}{6}\frac{2}{5}P\{X=4\}\approx0.036\,9$$

$$P\{X=6\}=\frac{4}{6}\frac{1}{6}P\{X=5\}\approx0.004\,1$$

■

利用计算二项分布函数的式(6-3)的递推公式很容易编写计算机程序．为计算 $P\{X\leqslant i\}$，程序应该先计算 $P\{X=i\}$，接下来应用递推式计算 $P\{X=i-1\}$，$P\{X=i-2\}$，等等．在网页上有此程序．运用此程序，输入二项参数 n,p,i，可以计算以 (n,p) 为参数的二项随机变量小于等于 i 的概率．

历史注记

成功的概率为 p 的独立试验最先由瑞士数学家雅克·伯努利(Jacques Bernoulli，1654—1705)所研究．在他的《Ars Conjectandi》(猜想的艺术，这本书在他去世后 8 年由他的侄子 Nicholas 于 1713 年整理出版)一书中，他证明了当试验的次数很大时，试验成功的频率以近似 1 的概率接近于 p．

雅克·伯努利来自最著名的数学家族中的第一代，三代一共有 8～12 个伯努利，他们对概率论、统计学和数学做出了奠基性的贡献．人们遇到的一个困难是不知道他们家中人口的确切数目，因为他们都叫伯努利．(例如，雅克的哥哥 Jean 的两个儿子分别叫 Jacques，Jean．)另外一个困难是几个伯努利在不同的地方有着不同的名字，如这里的 Jacques(有时写成 Jaques)也称为 Jakob(有时写成 Jacob)和 James，其实都是 Jacques Bernoulli．不管他们究竟有几个人，他们的贡献是巨大的．像乡村音乐一样，伯努利的家族是那个时代的伟大的数学之家．

148

例 6i 设 X 是参数为 $n=100$，$p=0.75$ 的二项随机变量，试求 $P\{X=70\}$ 与 $P\{X\leqslant 70\}$．

解 答案在如下网页上给出．

```
Binomial Distribution

Enter Value For p  .75          Start
Enter Value For n  100
Enter Value For i  70           Quit

Probability (Number of Successes = i) = .04575381
Probability (Number of Successes < = i) = .14954105
```

■

4.7 泊松随机变量

取值于 0,1,2,…的随机变量 X 称为以 λ 为参数的泊松随机变量(Poisson random variable)，如果对某一 $\lambda>0$ 有

$$p(i)=P\{X=i\}=e^{-\lambda}\frac{\lambda^i}{i!}\qquad i=0,1,2,\cdots\tag{7-1}$$

式(7-1)定义了一个概率质量函数，因为

$$\sum_{i=0}^{\infty}p(i)=e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}=e^{-\lambda}e^{\lambda}=1$$

泊松概率分布是泊松(S. D. Poisson)在他所著的关于概率论在诉讼、刑事审讯等方面的应用的书中引入的．这本书于 1837 年出版，书名为《Recherchés sur la probabilité des juge-

ments en matière criminelle et en matière civile》(概率论在刑事与民事诉讼方面应用的研究).

泊松随机变量在各种领域中有着极为广泛的应用，这是由于当 n 充分大、p 充分小而使得 np 保持适当的大小时，泊松分布可以用来作为以 (n,p) 为参数的二项随机变量的逼近. 为说明这一点，设 X 是参数为 (n,p) 的二项随机变量，并设 $\lambda=np$，则

149

$$P\{X=i\}=\frac{n!}{(n-i)!\ i!}p^i(1-p)^{n-i}=\frac{n!}{(n-i)!\ i!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i}$$
$$=\frac{n(n-1)\cdots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}$$

由于对充分大的 n 与适当的 λ，我们，有

$$\left(1-\frac{\lambda}{n}\right)^n\approx e^{-\lambda}\qquad\frac{n(n-1)\cdots(n-i+1)}{n^i}\approx 1\qquad\left(1-\frac{\lambda}{n}\right)^i\approx 1$$

因此对充分大的 n 与适当的 λ，有

$$P\{X=i\}\approx e^{-\lambda}\frac{\lambda^i}{i!}$$

换言之，如果进行 n 次独立重复试验，每次试验成功的概率为 p，那么当 n 充分大、p 充分小使得 np 保持适当的大小时，成功的次数近似于参数为 $\lambda=np$ 的泊松随机变量. 这个常数 λ（以后将证明它等于成功次数的期望值)通常根据经验来确定.

下面列出通常认为服从泊松概率定律[即满足式(7-1)]的随机变量的几个例子.

1. 一本书中某一页(或某几页)上印刷错误的个数.
2. 某地区居民中活到 100 岁的人数.
3. 一天中拨错号的电话呼叫次数.
4. 某商店在一天内卖出的动物饼干的包数.
5. 在某一天中进入某邮局的人数.
6. 在一年期间美国联邦法院出现的空缺数.
7. 在一个固定时期内，从某放射性物质中放射出的 α 粒子的数目.

上述每一个例子以及许多其他随机变量同样可以近似地看作泊松随机变量，这是因为泊松随机变量可近似地代替二项随机变量. 例如，可以假定书上某一页的每个字母印错的概率为 p（p 很小)，那么在这一页上印错的字母个数近似地为以 $\lambda=np$ 为参数的泊松随机变量，其中 n 为这页书上的字母总数. 类似地，可以假定某地区的每个人以某个小概率能活到 100 岁. 还有，可以设想进入商店的每个人会以某个小概率购买一包动物饼干，等等.

例 7a 假设书的某一页上印刷错误的个数服从以 $\lambda=\frac{1}{2}$ 为参数的泊松分布，试求在这一页上至少有一处印刷错误的概率.

解 令 X 表示这一页上印刷错误的个数，我们有

$$P\{X\geqslant 1\}=1-P\{X=0\}=1-e^{-\frac{1}{2}}\approx 0.393$$ ■

150

例 7b 设某个机器生产的产品是次品的概率为 0.1，试求 10 件样品中至多 1 件次品的概率.

解 所求的概率为

$$\binom{10}{0}(0.1)^0(0.9)^{10}+\binom{10}{1}(0.1)^1(0.9)^9=0.736\,1$$

运用泊松近似，所得的值为

$$e^{-1}+e^{-1}\approx 0.735\,8$$ ■

例 7c 考虑这样一个试验：记录 1 克放射性物质在 1 秒钟内放射出的 α 粒子的数目. 如果从过去的经验得知，这个数目平均为 3.2. 问对于放射出不超过 2 个 α 粒子的概率而言，较好的近似值是多少?

解 设想这 1 克放射性物质是由 n 个原子组成的(n 相当大)，每个原子在所考虑的 1 秒钟内衰变并放射出一个 α 粒子的概率为 $3.2/n$. 于是我们看到，作为十分接近的近似，放射出的 α 粒子的个数是以 $\lambda=3.2$ 为参数的泊松随机变量. 因此，所求的概率为

$$P\{X\leqslant 2\}=e^{-3.2}+3.2e^{-3.2}+\frac{(3.2)^2}{2}e^{-3.2}\approx 0.3799$$ ■

在计算参数为 λ 的泊松随机变量的期望值与方差之前，考虑以前将这个随机变量近似为参数为(n,p)的二项随机变量，其中 n 充分大、p 足够小且 $\lambda=np$. 因为二项随机变量的数学期望为 $np=\lambda$，方差为 $np(1-p)=\lambda(1-p)\approx\lambda$(因为 p 很小)，所以我们可以看到泊松随机变量的期望值与方差均为 λ. 下面验证这个结论.

$$\begin{aligned}E[X] &= \sum_{i=0}^{\infty}\frac{ie^{-\lambda}\lambda^i}{i!}=\lambda\sum_{i=1}^{\infty}\frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!}\\ &=\lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} \qquad 令\ j=i-1\\ &=\lambda \qquad 因为\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}=e^{\lambda}\end{aligned}$$

151

泊松随机变量 X 的期望值确实等于参数λ. 为计算方差，可以先计算 $E[X^2]$.

$$\begin{aligned}E[X^2] &= \sum_{i=0}^{\infty}\frac{i^2e^{-\lambda}\lambda^i}{i!}=\lambda\sum_{i=1}^{\infty}\frac{ie^{-\lambda}\lambda^{i-1}}{(i-1)!}\\ &=\lambda\sum_{j=0}^{\infty}\frac{(j+1)e^{-\lambda}\lambda^j}{j!} \qquad 令\ j=i-1\\ &=\lambda\left[\sum_{j=0}^{\infty}\frac{je^{-\lambda}\lambda^j}{j!}+\sum_{j=0}^{\infty}\frac{e^{-\lambda}\lambda^j}{j!}\right]=\lambda(\lambda+1)\end{aligned}$$

其中上面最后一个等号成立是因为第一项等于参数为 λ 的泊松随机变量的期望值，第二项是这个随机变量的概率之和. 因为已知 $E[X]=\lambda$，所以我们可以得到

$$\mathrm{Var}(X)=E[X^2]-(E[X])^2=\lambda$$

泊松随机变量的期望值与方差均为 λ.

对 n 个独立试验，每次成功的概率为 p，当 n 充分大、p 足够小时，试验成功的次数很好地近似于参数为 np 的泊松分布. 事实上，当试验不相互独立或依赖性很弱时，泊松分布仍是很好的近似. 例如，考虑配对问题(第 2 章例 5 m)，n 个人随机选择帽子. 可以把选对帽子的人数看成是 n 个独立试验成功的次数. 如果第 i 个人选对自己的帽子，我们认为试验 i 成功，$i=1,\cdots,n$. 定义事件 $E_i(i=1,\cdots,n)$为

$$E_i=\{试验\ i\ 成功\}$$

很容易看到

$$P\{E_i\}=\frac{1}{n}与 P\{E_i|E_j\}=\frac{1}{n-1},j\neq i$$

显然，事件 $E_i(i=1,\cdots,n)$不是相互独立的，当 n 很大时，可以认为它们之间的依赖性很弱. 基于此，将试验成功次数近似为参数为 $n\times 1/n=1$ 的泊松分布是合理的，事实上，第 2 章例 5 m 中已经验证过了.

152

再说明一下，当试验的依赖性很弱时，泊松分布仍是很好的近似. 重新看第 2 章例 5 i. 在这个例子中，假定 n 个人等可能地在 365 天的任一天过生日，问题是求 n 个独立的人的任意组合都有不同生日的概率. 用组合的方法计算此概率，当 $n=23$ 时，这个概率小于$\frac{1}{2}$.

我们可以将上面的概率近似为泊松分布. 试验 i 与试验 j 有$\binom{n}{2}$种组合，如果第 i 人与第 j 人生日相同，就说明试验 i，j 成功，E_{ij}表示试验 i，j 成功的事件，而$\binom{n}{2}$个事件 $E_{ij}(1\leqslant i<j\leqslant n)$不是相互独立的(见理论练习 21)，但它们的依赖性很弱. (事实上，这些事件是两两独立的，即事件 E_{ij} 与 E_{kl} 是相互独立的，参见理论练习 21.)由于 $P(E_{ij})=1/365$，所以将试验成功的次数近似成参数为$\binom{n}{2}\Big/365=n(n-1)/730$ 的泊松分布. 因此，

$$P\{\text{任何两人不同一天过生日}\}=P\{0\text{ 次成功}\}\approx\exp\left\{\frac{-n(n-1)}{730}\right\}$$

求使此概率小于$\frac{1}{2}$的最小整数 n，即

$$\exp\left\{\frac{-n(n-1)}{730}\right\}\leqslant\frac{1}{2}$$

或等价地

$$\exp\left\{\frac{n(n-1)}{730}\right\}\geqslant 2$$

上式两边取对数，可以得到

$$n(n-1)\geqslant 730\log 2\approx 505.997$$

解出 $n=23$，与第 2 章例 5i 的结果一致.

假设考虑 n 个人中任何三个都不在同一天过生日的概率. 这是一个困难的组合问题，很容易得到一个很好的近似. 假定我们开始一个试验，其中$\binom{n}{3}$中的 3 个试验 i,j,k 一组，$1\leqslant i<j<k\leqslant n$,如果第 i,j,k 人生日相同，就说明试验 i,j,k 成功. 如前，我们可以推出试验成功的次数用泊松随机变量来近似，参数为

153

$$\binom{n}{3}P\{i,j,k\text{ 有相同的生日}\}=\binom{n}{3}\left(\frac{1}{365}\right)^2=\frac{n(n-1)(n-2)}{6\times(365)^2}$$

因此

$$P\{\text{任何三个人不同一天过生日}\}\approx\exp\left\{\frac{-n(n-1)(n-2)}{799\ 350}\right\}$$

当 n 满足

$$n(n-1)(n-2)\geqslant 799\ 350\log 2\approx 554\ 067.1$$

时，这个概率小于$\frac{1}{2}$. 解出 $n\geqslant 84$. 因此，在大于等于 84 人的集合中至少有 3 个人生日相同的近似概率将会超过$\frac{1}{2}$.

泊松概率分布应用的其他例子还表现在“事件”在确定的时间点列上发生的情形. 这种“事件”的例子有：发生一次地震，一个人独自进入某特定的地方(如银行、邮局、加油站等)，爆发一场战争，等等. 假定这样的“事件”实际上发生在一列(随机)时间点上，并存在某个正的常数 λ 使如下条件成立：

1. 在任意长度为 h 的时间间隔内，正好发生一个"事件"的概率等于 $\lambda h+o(h)$，其中 $o(h)$ 表示任何满足 $\lim\limits_{h\to 0}f(h)/h=0$ 的函数 $f(h)$. [例如 $f(h)=h^2$ 是 $o(h)$，而 $f(h)=h$ 不是 $o(h)$.]

2. 在任意长度为 h 的时间间隔内发生两个或更多个"事件"的概率等于 $o(h)$.

3. 对于任意整数 $n,j_1,j_2,\cdots,j_n$ 以及任意 n 个互不相交的时间间隔，若以 E_i 表示在第 i 个时间间隔内上述"事件"正好发生 j_i 个这一事件，则 $E_1,E_2,\cdots,E_n$ 是相互独立的.

粗略地说，条件 1 与 2 说明当 h 比较小时，在长度为 h 的时间间隔内正好发生一个"事件"的概率等于 λh 加上某个比 h 更小的量，而两个或更多个事件发生的概率小于 h. 条件 3 说明在某一个时间间隔内出现的事件对另一个与其不相交的时间间隔内将要发生的事件(从概率上说)没有影响.

在条件 1，2 与 3 成立的假定下，我们来证明，在任意长度为 t 的时间间隔内，"事件"发生的个数是以 λt 为参数的泊松随机变量. 为此，我们考虑区间$[0,t]$，并以 $N(t)$表示这个区间内事件发生的个数. 为求出 $P\{N(t)=k\}$的表达式，先将区间$[0,t]$等分为 n 个互不相交且长度为 t/n 的子区间(见图 4-6). 154



图 4-6

我们有

$$P\{N(t)=k\}=P\{n\text{ 个子区间中的 }k\text{ 个正好各含一个事件,而其余 }n-k\text{ 个子区间各含 0 个事件}\} \quad (7\text{-}2)$$
$$+P\{N(t)=k\text{ 且至少 1 个子区间含有两个或更多个事件}\}$$

这是因为式(7-2)左边的事件$\{N(t)=k\}$显然等于右边的两个互不相容事件之并. 以 A 与 B 分别表示式(7-2)右边的两个互不相容事件，便得到

$$\begin{aligned}P(B)&\leqslant P\{\text{至少一个子区间含有两个或更多个事件}\}\\&=P\left(\bigcup_{i=1}^{n}\{\text{第 }i\text{ 个子区间含有两个或更多个事件}\}\right)\\&\leqslant\sum_{i=1}^{n}P\{\text{第 }i\text{ 个子区间含有两个或更多个事件}\}\qquad\text{由布尔不等式}\\&=\sum_{i=1}^{n}o\left(\frac{t}{n}\right)\qquad\text{由条件 2}\\&=no\left(\frac{t}{n}\right)=t\left[\frac{o(t/n)}{t/n}\right]\end{aligned}$$

现在，对任何 t，当 $n\to\infty$时 $t/n\to 0$，从而由 $o(h)$的定义可得当 $n\to\infty$时，$o(t/n)/(t/n)\to 0$. 因此，当 $n\to\infty$时，

$$P(B)\to 0 \quad (7\text{-}3)$$

另一方面，由于条件 1 与 2 蕴涵着[⊖]

$$P\{\text{在长度为 }h\text{ 的区间内有 0 个事件发生}\}=1-[\lambda h+o(h)+o(h)]=1-\lambda h-o(h)$$ 155

由独立性条件 3 可见

$$P(A)=P\{n\text{ 个子区间中的 }k\text{ 个正好各含一个事件，而其余 }n-k\text{ 个子区间各含 0 个事件}\}$$

⊖ 两个同为 $o(h)$的函数之和仍为 $o(h)$. 这是因为，若$\lim\limits_{h\to 0}f(h)/h=\lim\limits_{h\to 0}g(h)/h=0$，则$\lim\limits_{h\to 0}[f(h)+g(h)]/h=0$.

$$=\binom{n}{k}\left[\frac{\lambda t}{n}+o\left(\frac{t}{n}\right)\right]^{k}\left[1-\left(\frac{\lambda t}{n}\right)-o\left(\frac{t}{n}\right)\right]^{n-k}$$

但是，当 $n\to\infty$ 时，

$$n\left[\frac{\lambda t}{n}+o\left(\frac{t}{n}\right)\right]=\lambda t+t\left[\frac{o(t/n)}{t/n}\right]\to\lambda t$$

采用与证明二项随机变量的泊松近似相同的方法可证，当 $n\to\infty$ 时，

$$P(A)\to e^{-\lambda t}\frac{(\lambda t)^{k}}{k!} \tag{7-4}$$

至此，由式(7-2)，(7-3)与(7-4)，当 $n\to\infty$ 时，我们得到

$$P\{N(t)=k\}=e^{-\lambda t}\frac{(\lambda t)^{k}}{k!}\qquad k=0,1,2,\cdots \tag{7-5}$$

这样，如果事件满足条件 1,2 与 3，则在任何固定的长度为 t 的时间区间内，事件发生的个数是以 λt 为参数的泊松随机变量，这时我们称事件是按速度为 λ 的泊松过程而发生的. λ 可以认为是单位时间内事件发生的速度，它是根据经验确定的常数.

上述讨论解释了为什么泊松随机变量通常可作为下面各种现象的很好的近似.

1. 在某固定时间间隔内发生地震的次数.
2. 每年爆发的战争次数.
3. 在固定周期内从一个热阴极发射出的电子数.
4. 某人寿保险公司的保险客户在某一时间间隔内死亡的人数.

例 7d 设在美国西部发生地震符合条件 1,2,3，其中时间以周为单位，$\lambda=2$.（即假定以每周 2 次的频率按三个条件要求的方式发生地震.）

(a) 求在相邻两周内至少发生 3 次地震的概率.

156

(b) 求从现在起到下一次地震发生所需时间的概率分布.

解 (a) 由式(7-5)可得

$$P\{N(2)\geqslant 3\}=1-P\{N(2)=0\}-P\{N(2)=1\}-P\{N(2)=2\}=1-e^{-4}-4e^{-4}-\frac{4^2}{2}e^{-4}=1-13e^{-4}$$

(b) 以 X 表示直到下一次地震发生所需的时间(单位为周). 当且仅当从现在起 t 周内无地震时，X 比 t 大，故由式(7-5)可得

$$P\{X>t\}=P\{N(t)=0\}=e^{-\lambda t}$$

从而随机变量 X 的概率分布函数 F 为

$$F(t)=P\{X\leqslant t\}=1-P\{X>t\}=1-e^{-\lambda t}=1-e^{-2t}$$ ■

计算泊松分布函数

若 X 是参数为 λ 的泊松随机变量，则

$$\frac{P\{X=i+1\}}{P\{X=i\}}=\frac{e^{-\lambda}\lambda^{i+1}/(i+1)!}{e^{-\lambda}\lambda^{i}/i!}=\frac{\lambda}{i+1} \tag{7-6}$$

因为 $P\{X=0\}=e^{-\lambda}$，利用式(7-6)可得如下递推公式：

$$P\{X=1\}=\lambda P\{X=0\}$$

$$P\{X=2\}=\frac{\lambda}{2}P\{X=1\}$$

$$\vdots$$

$$P\{X=i+1\}=\frac{\lambda}{i+1}P\{X=i\}$$

网站中包含了使用式(7-6)来计算泊松概率的程序.

例 7e (a) 随机变量 X 服从参数为 100 的泊松分布，求 $P\{X\leqslant 90\}$.

(b) 随机变量 Y 服从参数为 1 000 的泊松分布，求 $P\{Y\leqslant 1\,075\}$.

解 从网站可得

(a) $P\{X\leqslant 90\}\approx 0.171\,4$.

(b) $P\{Y\leqslant 1\,075\}\approx 0.989\,4$. ■ 157

4.8 其他离散型概率分布

4.8.1 几何随机变量

重复进行一个每次成功的概率为 $p(0<p<1)$的独立试验到出现一次成功为止. 若设 X 为所需的试验次数，则

$$P\{X=n\}=(1-p)^{n-1}p,\qquad n=1,2,\cdots \tag{8-1}$$

式(8-1)成立是因为 X 等于 n 的充分必要条件是前 $n-1$ 次试验都失败而第 n 次试验成功，再考虑各次试验结果相互独立就可得到式(8-1).

因为

$$\sum_{n=1}^{\infty}P\{X=n\}=p\sum_{n=1}^{\infty}(1-p)^{n-1}=\frac{p}{1-(1-p)}=1$$

所以，最终取得成功的概率为 1. 概率质量函数由式(8-1)给出的随机变量 X 称为以 p 为参数的几何随机变量(geometric random variable).

例 8a 箱中有 N 个白球与 M 个黑球，每次随机取一个球，直到取出黑球为止. 如果每取出一球后立即放回，再取出一个球，试求下列概率.

(a) 正好需要取 n 次.

(b) 至少需要取 k 次.

解 设 X 表示取到黑球所需的次数，则 X 满足式(8-1)，其中 $p=M/(M+N)$. 于是有

(a) $P\{X=n\}=\left(\frac{N}{M+N}\right)^{n-1}\frac{M}{M+N}=\frac{MN^{n-1}}{(M+N)^n}$

(b) $P\{X\geqslant k\}=\frac{M}{M+N}\sum_{n=k}^{\infty}\left(\frac{N}{M+N}\right)^{n-1}=\left(\frac{M}{M+N}\right)\left(\frac{N}{M+N}\right)^{k-1}\Big/\left[1-\frac{N}{M+N}\right]=\left(\frac{N}{M+N}\right)^{k-1}$

当然，(b)也可以直接计算，这是因为至少需 k 次试验才能取得成功的概率等于前 $k-1$ 次试验都失败的概率. 即，对几何随机变量而言

$$P\{X\geqslant k\}=(1-p)^{k-1}$$

■ 158

例 8b 计算几何随机变量的数学期望.

解 令 $q=1-p$，我们有

$$E[X]=\sum_{n=1}^{\infty}nq^{n-1}p=p\sum_{n=0}^{\infty}\frac{\mathrm{d}}{\mathrm{d}q}(q^n)=p\frac{\mathrm{d}}{\mathrm{d}q}\left(\sum_{n=0}^{\infty}q^n\right)=p\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{1}{1-q}\right)=\frac{p}{(1-q)^2}=\frac{1}{p}$$

换言之，独立试验每次成功的概率为 p，进行到出现一次成功为止，则所需的试验期望次

数为 $1/p$. 例如，掷一个均匀的骰子得到点数 1 需掷 6 次. ■

例 8c 计算几何随机变量的方差.

解 为计算 $\mathrm{Var}(X)$，先计算 $E[X^2]$. 令 $q=1-p$，

$$E[X^2]=\sum_{n=1}^{\infty}n^2q^{n-1}p=p\sum_{n=1}^{\infty}\frac{\mathrm{d}}{\mathrm{d}q}(nq^n)=p\frac{\mathrm{d}}{\mathrm{d}q}\left(\sum_{n=1}^{\infty}nq^n\right)=p\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{q}{1-q}E[X]\right)$$

$$=p\frac{\mathrm{d}}{\mathrm{d}q}[q(1-q)^{-2}]=p\left[\frac{1}{p^2}+\frac{2(1-p)}{p^3}\right]=\frac{2}{p^2}-\frac{1}{p}$$

因为 $E[X]=1/p$，所以

$$\mathrm{Var}(X)=\frac{1-p}{p^2}$$ ■

159

4.8.2 负二项随机变量

进行独立重复试验，每次成功的概率为 p，$0<p<1$，直到出现 r 次成功为止. 设 X 表示所需的试验次数，则

$$P\{X=n\}=\binom{n-1}{r-1}p^r(1-p)^{n-r},n=r,r+1,\cdots \tag{8-2}$$

式(8-2)成立是因为，为使第 r 次成功发生在第 n 次试验，必须在前 $n-1$ 次试验中取得 $r-1$ 次成功，而且第 n 次试验也一定成功. 前一事件的概率为

$$\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}$$

后一事件的概率为 p；再由独立性即得式(8-2). 为证实 r 次成功最终实现，我们可以利用解析方法证明

$$\sum_{n=r}^{\infty}P\{X=n\}=\sum_{n=r}^{\infty}\binom{n-1}{r-1}p^r(1-p)^{n-r}=1 \tag{8-3}$$

也可以用如下概率的方法论证：将取得 r 次成功所需的试验次数表示为 $Y_1+Y_2+\cdots+Y_r$，其中 Y_1 为取得第一次成功所需的试验次数，Y_2 为第一次成功后要取得第二次成功需再做的试验次数，Y_3 表示为取得第三次成功需再做的试验次数，如此等等. 由于重复试验是相互独立的，而且每次试验成功的概率相同，因此 $Y_1,Y_2,\cdots,Y_r$ 是几何随机变量. 于是，每个 Y_i 以概率 1 是有限的，从而它们的和 $\sum_{i=1}^{r}Y_i$ 也一定是有限的，这就得到了式(8-3).

概率质量函数由式(8-2)给出的随机变量 X 称为以 (r,p) 为参数的**负二项随机变量**(negative binomial random variable). 易见，几何随机变量是以 $(1,p)$ 为参数的负二项随机变量.

在下一个例子中，我们运用负二项随机变量给出分赌注问题的另一解法.

例 8d 进行一独立重复试验，每次结果成功的概率为 p. 问在 m 次失败之前取得 r 次成功的概率是多少?

解 我们指出，为使 r 次成功发生在 m 次失败之前，当且仅当第 r 次成功不在第 $r+m-1$ 次试验之后发生. 这是因为，如果第 r 次成功在第 $r+m-1$ 次试验或在此试验之前获得，那么第 r 次成功必在第 m 次失败之前，反之亦然. 因此，由式(8-2)可知所求的概率为

160

$$\sum_{n=r}^{r+m-1}\binom{n-1}{r-1}p^r(1-p)^{n-r}$$ ■

例 8e（巴拿赫火柴问题）　一位吸烟的数学家总带着两盒火柴，每盒有 N 根. 一盒放在左边衣袋中，另一盒放在右边衣袋中. 每当他需用一根火柴时，就等可能地从两个衣袋中取用. 试问他第一次发现一盒是空的而另一盒正好还有 $k(k=0,1,\cdots,N)$根的概率是多少？

解　设 E 表示这位数学家右边衣袋中的火柴盒是空的而左边衣袋中的火柴盒还有 k 根火柴这一事件. 由于此事件当且仅当他第 $N+1$ 次取出右边衣袋中的火柴盒且正好是他第 $N+1+N-k$ 次用火柴时才发生，故由式(8-2)(取 $p=\frac{1}{2}$，$r=N+1$，$n=2N-k+1$)可得

$$P(E)=\binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k+1}$$

因为他发现左边衣袋中的火柴盒是空的而右边衣袋中的火柴盒还有 k 根的概率与上面的概率相同，所以所求的概率为

$$2P(E)=\binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k}$$ ■

例 8f　计算以(r,p)为参数的负二项随机变量的数学期望和方差.

解

$$\begin{aligned}
E[X^k] &= \sum_{n=r}^{\infty} n^k \binom{n-1}{r-1} p^r (1-p)^{n-r} \\
&= \frac{r}{p}\sum_{n=r}^{\infty} n^{k-1}\binom{n}{r} p^{r+1}(1-p)^{n-r} \quad \text{因为 } n\binom{n-1}{r-1}=r\binom{n}{r} \\
&= \frac{r}{p}\sum_{m=r+1}^{\infty}(m-1)^{k-1}\binom{m-1}{r}p^{r+1}(1-p)^{m-(r+1)} \quad \text{令 } m=n+1 \\
&= \frac{r}{p}E[(Y-1)^{k-1}]
\end{aligned}$$

其中 Y 是参数为$(r+1,p)$的负二项随机变量，在上式中令 $k=1$ 得到

$$E[X]=\frac{r}{p}$$

在上式中令 $k=2$ 得到

$$E[X^2]=\frac{r}{p}E[Y-1]=\frac{r}{p}\left(\frac{r+1}{p}-1\right)$$

因此

$$\mathrm{Var}(X)=\frac{r}{p}\left(\frac{r+1}{p}-1\right)-\left(\frac{r}{p}\right)^2=\frac{r(1-p)}{p^2}$$ ■

从例 8f 可以看到，如果独立试验每次成功的概率为 p，则 r 次成功的期望值与方差分别为 r/p 与 $r(1-p)/p^2$.

由于几何随机变量正好是以 $r=1$ 为参数的负二项随机变量，从上例可知参数为 p 的几何随机变量的方差等于$(1-p)/p^2$，这和例 8c 的结果一致.

例 8g　掷一个骰子直到出现 4 次点数为 1，试求其数学期望与方差.

解　由题意知，这是参数为 $r=4$，$p=\frac{1}{6}$的负二项随机变量，因此

$$E[X]=24$$

$$\mathrm{Var}(X)=\frac{4\left(\frac{5}{6}\right)}{\left(\frac{1}{6}\right)^2}=120$$

■

4.8.3 超几何随机变量

假设箱中有 N 个球，其中 m 个白球、$N-m$ 个黑球，不放回地从箱中随机取出 n 个球. 设 X 表示取出的白球数，则

$$P\{X=i\}=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \qquad i=0,1,\cdots,n \tag{8-4}$$

概率质量函数由式(8-4)给出的随机变量 X 称为以 n,N,m 为参数的超几何随机变量(hypergeometric random variable).

注 虽然我们已经写出了 i 从 0 到 n 时的超几何随机变量的概率质量函数，除 i 满足不等式 $n-(N-m)\leqslant i\leqslant \min(n,m)$ 外，$P\{X=i\}$ 实际上等于 0. 然而，式(8-4)总是成立的，因为当 $k<0$ 或 $r<k$ 时，我们约定 $\binom{r}{k}$ 等于 0.

162

例 8h 栖居于某地区的动物个数 N 是未知的. 为得到种群大小的有关信息，生态学家们常常进行如下试验：他们先在这个地区捉一些动物，比如说 m 个，标上某种记号后放掉它们. 过一段时间，当那些标过记号的动物充分地散布在整个地区后，再捉一批，比如说 n 个，设 X 为第二次捉的 n 个动物中标过记号的动物个数. 如果假设在两次捕捉期间这个地区动物的总数没有变化，而且每捉一个动物都是在剩下的所有动物中等可能地选择，则 X 是一个超几何随机变量，满足

$$P\{X=i\}=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}\equiv P_i(N)$$

现设 X 的观察值为 i，由于 $P_i(N)$ 是在这个地区事实上总共有 N 个动物的条件下所观察到的事件的概率，因此看来使 $P_i(N)$ 达到最大值的 N 应当是动物总数 N 的合乎情理的估计. 这样取的估计量称为**最大似然估计**(maximum likelihood estimate). (这种估计方法的其他例子见理论练习 13 与 18.)

求 $P_i(N)$ 最大值的最简单方法是，首先注意到

$$\frac{P_i(N)}{P_i(N-1)}=\frac{(N-m)(N-n)}{N(N-m-n+i)}$$

为使上式比值大于 1，当且仅当

$$(N-m)(N-n)\geqslant N(N-m-n+i)$$

或等价地，当且仅当

$$N\leqslant\frac{mn}{i}$$

可见，$P_i(N)$ 对 N 而言先上升、后下降，且在不超过 $\frac{mn}{i}$ 的最大整数处达到其最大值. 这个最大整数就是 N 的最大似然估计. 例如，假定第一次捉了 $m=50$ 只动物，标上记号后放掉，第二次

捉了 $n=40$ 只动物，其中已标上记号的有 $i=4$ 只，那么，我们就可以估计出这个地区大约有 500 只动物.（上述估计还可以这样求得：在这个地区内，标上记号的动物所占的比例 m/N 近似地等于在第二次捉的动物中做过标号的动物所占的比例 i/n.） ■

163

例 8i 某采购员购买一种 10 个一包的电器元件，他的方法是从一包中随机地抽查 3 个，如果这 3 个元件都是好的才买下这一包. 假定含有 4 个坏元件的包数占 30%，而其余 70% 每包只含 1 个坏元件. 试问这个采购员拒绝购买的包数占多大的比例?

解 设 A 表示采购员买下一包这一事件，则

$$P(A)=P(A|\text{这包中含有 4 个次品})\frac{3}{10}+P(A|\text{这包中含有 1 个次品})\frac{7}{10}$$

$$=\frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}}\left(\frac{3}{10}\right)+\frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}}\left(\frac{7}{10}\right)=\frac{54}{100}$$ ■

从而有 46% 的包被采购员拒绝.

从 N 个球中不放回地取出 n 个球，其中白球占 $p=m/N$，选出的白球数是超几何随机变量. 当 m 和 N 远远大于 n 时，有放回和无放回没有很大差别. 因为当 m 和 N 很大时，不管已经取出了多少个球，再次取出白球的概率仍近似等于 p. 换句话说，可以直观地看出，当 m 和 N 很大时，X 的概率质量函数近似为参数为 n 和 p 的二项随机变量. 为验证这一点，注意若 X 是超几何随机变量，对 $i\leqslant n$，

$$P\{X=i\}=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}=\frac{m!}{(m-i)!\ i!}\frac{(N-m)!}{(N-m-n+i)!\ (n-i)!}\frac{(N-n)!\ n!}{N!}$$

$$=\binom{n}{i}\frac{m}{N}\frac{m-1}{N-1}\cdots\frac{m-i+1}{N-i+1}\frac{N-m}{N-i}\frac{N-m-1}{N-i-1}\cdots\frac{N-m-(n-i-1)}{N-i-(n-i-1)}$$

$$\approx\binom{n}{i}p^i(1-p)^{n-i}\qquad \text{其中 } p=m/N,\ m \text{ 和 } N \text{ 远远大于 } n \text{ 与 } i$$

164

例 8j 计算以 (n,N,m) 为参数的超几何随机变量 X 的数学期望与方差.

解

$$E[X^k]=\sum_{i=0}^{n}i^kP\{X=i\}=\sum_{i=1}^{n}i^k\binom{m}{i}\binom{N-m}{n-i}\Big/\binom{N}{n}$$

由恒等式

$$i\binom{m}{i}=m\binom{m-1}{i-1}\quad \text{与}\quad n\binom{N}{n}=N\binom{N-1}{n-1}$$

可以得到

$$E[X^k]=\frac{nm}{N}\sum_{i=1}^{n}i^{k-1}\binom{m-1}{i-1}\binom{N-m}{n-i}\Big/\binom{N-1}{n-1}$$

$$=\frac{nm}{N}\sum_{j=0}^{n-1}(j+1)^{k-1}\binom{m-1}{j}\binom{N-m}{n-1-j}\Big/\binom{N-1}{n-1}$$

$$=\frac{nm}{N}E[(Y+1)^{k-1}]$$

其中 Y 是以 $(n-1, N-1, m-1)$ 为参数的超几何随机变量. 令 $k=1$，得到

$$E[X]=\frac{nm}{N}$$

因此，如果从 N 个球的集合中随机取出 n 个球，其中 m 个白球，则取到的白球的期望值为 nm/N.

令 $k=2$，得到

$$E[X^2]=\frac{nm}{N}E[Y+1]=\frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1}+1\right]$$

其中最后一个等号是用前面的公式来计算超几何随机变量 Y 的期望值.

又因为 $E[X]=\frac{nm}{N}$，可以推出

$$\mathrm{Var}(X)=\frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1}+1-\frac{nm}{N}\right] \tag{8-5}$$

令 $p=m/N$ 表示白球所占的比例，对式(8-5)作一变换可得

$$\mathrm{Var}(X)=\frac{N-n}{N-1}np(1-p) \tag{8-6}$$

■

165

注 例 8j 说明若从 N 个球中不放回地取出 n 个球，p 是白球所占的比例，则取出的白球的数学期望值为 np. 另外，若 N 远远大于 n[也即 $(N-n)/(N-1)$ 近似等于 1]，则

$$\mathrm{Var}(X)\approx np(1-p)$$

换言之，$E[X]$ 与有放回地抽球的结果是相同的[白球数服从二项分布，参数为 n, p]，而且，如果箱中球很多，则 $\mathrm{Var}(X)$ 近似等于有放回抽球情形的方差. 这是因为当箱中球的数量很多时，我们猜测所抽出的白球数近似地有二项随机变量的质量函数.

4.8.4 ζ(Zipf)分布

一个随机变量称为 ζ 分布[有时称为 Zipf 分布]，如果对某个 $\alpha>0$，其概率质量函数为

$$P\{X=k\}=\frac{C}{k^{\alpha+1}}, k=1,2,\cdots$$

由于上述概率总和为 1，故

$$C=\left[\sum_{k=1}^{\infty}\left(\frac{1}{k}\right)^{\alpha+1}\right]^{-1}$$

ζ 分布这个名称来源于数学学科，函数

$$\zeta(s)=1+\left(\frac{1}{2}\right)^s+\left(\frac{1}{3}\right)^s+\cdots+\left(\frac{1}{k}\right)^s+\cdots$$

称为黎曼 ζ 函数[源于德国数学家黎曼(G. F. B. Riemann)].

意大利经济学家佩瑞多(Pareto)曾用 ζ 分布来描述某个国家家庭收入的分布. 另外，G. K. Zipf 把这一分布运用到更广泛的领域，从而推广了它的用途.

4.9 累积分布函数的性质

对随机变量 X 的分布函数 F，$F(b)$ 表示随机变量 X 取小于或者等于 b 的某个值的概率，累积分布函数(cumulative distribution function, c. d. f.)F 有如下一些性质：

1. F 是非降函数；即若 $a<b$，则 $F(a)\leqslant F(b)$.
2. $\lim\limits_{b\to\infty}F(b)=1$.
3. $\lim\limits_{b\to-\infty}F(b)=0$.

166

4. F 右连续，即对任意 b 以及任意递减序列 b_n（其中 $n\geqslant 1$，b_n 收敛于 b），$\lim\limits_{n\to\infty}F(b_n)=F(b)$. 换言之，

$$\lim_{b\to b_0-} F(b)=F(b_0)$$

其中 $\lim\limits_{b\to b_0+}$ 表示在 $b>b_0$ 条件下，$b\to b_0$ 时所得到的极限.

性质 1 成立是因为在 4.1 节中，若 $a<b$，事件 $\{X\leqslant a\}$ 包含在事件 $\{X\leqslant b\}$ 之中，从而不会有更大的概率. 性质 2，3，4 可以从概率的连续性(2.6 节)推得. 例如，为证性质 2，我们指出当 $b_n\to\infty$ 时，事件序列 $\{X\leqslant b_n\}$ $(n\geqslant 1)$ 收敛于 $\{X<\infty\}$. 于是，由概率的连续性可得

$$\lim_{n\to\infty}P\{X\leqslant b_n\}=P\{x<\infty\}=1$$

这就证明了性质 2.

性质 3 的证明是类似的，留作习题. 为证性质 4，我们指出若 $b_n\to b$，则事件序列 $\{X\leqslant b_n\}$ 收敛于 $\{X\leqslant b\}$. 于是，由概率的连续性可得

$$\lim_{n}P\{X\leqslant b_n\}=P\{X\leqslant b\}$$

从而性质 4 得证.

有关 X 的一切概率问题都可以运用累积分布函数 F 来解答. 例如，对一切 $a<b$ 有

$$P\{a<X\leqslant b\}=F(b)-F(a) \tag{9-1}$$

式(9-1)也可以将事件 $\{X\leqslant b\}$ 写成互不相容的事件 $\{X\leqslant a\}$ 与 $\{a<X\leqslant b\}$ 之并而得到，即

$$\{X\leqslant b\}=\{X\leqslant a\}\cup\{a<X\leqslant b\}$$

从而

$$P\{X\leqslant b\}=P\{X\leqslant a\}+P\{a<X\leqslant b\}$$

于是式(9-1)得证.

如果要计算 X 严格小于 b 的概率，可以再次运用概率的连续性得到

$$P\{X<b\}=P\left(\lim_{n\to\infty}\left\{X\leqslant b-\frac{1}{n}\right\}\right)=\lim_{n\to\infty}P\left(X\leqslant b-\frac{1}{n}\right)=\lim_{n\to\infty}F\left(b-\frac{1}{n}\right)$$

注意 $P\{X<b\}$ 不一定等于 $F(b)$，因为 $F(b)$ 还包含了 X 等于 b 的概率.

167

例 9a 设随机变量 X 的分布函数为

$$F(x)=\begin{cases}0 & x<0\\[4pt] \dfrac{x}{2} & 0\leqslant x<1\\[4pt] \dfrac{2}{3} & 1\leqslant x<2\\[4pt] \dfrac{11}{12} & 2\leqslant x<3\\[4pt] 1 & 3\leqslant x\end{cases}$$

$F(x)$ 的图形如图 4-7 所示，试求

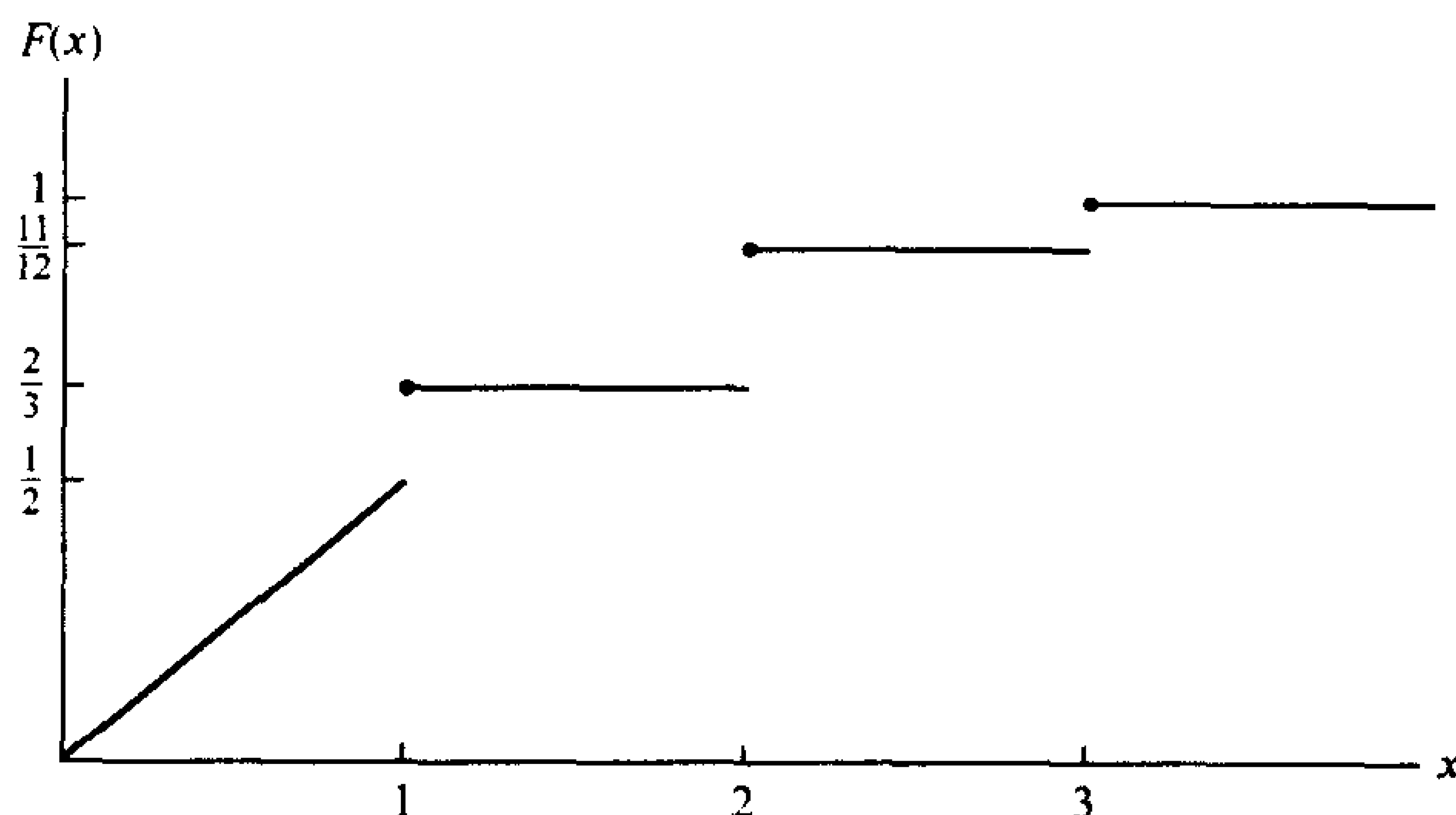
(a) $P\{X<3\}$，(b) $P\{X=1\}$，(c) $P\left\{X>\dfrac{1}{2}\right\}$，(d) $P\{2<X\leqslant 4\}$.

解 (a) $P\{X<3\}=\lim\limits_{n}P\left\{X\leqslant 3-\dfrac{1}{n}\right\}=\lim\limits_{n}F\left(3-\dfrac{1}{n}\right)=\dfrac{11}{12}$.

(b) $P\{X=1\}=P\{X\leqslant 1\}-P\{X<1\}=F(1)-\lim_n F\left(1-\frac{1}{n}\right)=\frac{2}{3}-\frac{1}{2}=\frac{1}{6}$.

(c) $P\left\{X>\frac{1}{2}\right\}=1-P\left\{X\leqslant\frac{1}{2}\right\}=1-F\left(\frac{1}{2}\right)=\frac{3}{4}$.

(d) $P\{2<X\leqslant 4\}=F(4)-F(2)=\frac{1}{12}$.



168

图 4-7 $F(x)$的图形 ■

小结

定义在试验结果上的实值函数称为随机变量.

若 X 是随机变量，则定义为

$$F(x)=P\{X\leqslant x\}$$

的函数 $F(x)$称为 X 的分布函数. 关于 X 的所有概率均可以通过 F 来表示.

可能取的值的集合为有限集或可数无限集的随机变量称为离散型随机变量. 若 X 是离散型随机变量，则函数

$$p(x)=P\{X=x\}$$

称为 X 的概率质量函数. 而且，$E[X]$定义为

$$E[X]=\sum_{x:p(x)>0}xp(x)$$

称为 X 的数学期望. $E[X]$通常也称为均值或 X 的期望值.

对函数 g，有如下有用的表示：

$$E[g(X)]=\sum_{x:p(x)>0}g(x)p(x)$$

随机变量 X 的方差，记为 $\text{Var}(X)$，定义为

$$\text{Var}(X)=E[(X-E[X])^2]$$

方差等于 X 与其期望值之差的平方取数学期望，它是 X 的各个可能值分散情况的一种度量. 一个有用的恒等式为

$$\text{Var}(X)=E[X^2]-(E[X])^2$$

$\sqrt{\text{Var}(X)}$称为 X 的标准差.

下面给出一些常用的离散型随机变量.

概率质量函数为

$$p(i)=\binom{n}{i}p^i(1-p)^{n-i} \qquad i=0,\cdots,n$$

的随机变量 X 称为参数为(n,p)的**二项随机变量**. 这样的随机变量表示 n 次独立试验中成功的次数，每次试验成功的概率为 p. 其期望和方差如下：

$$E[X]=np \qquad \mathrm{Var}(X)=np(1-p)$$

169

概率质量函数为

$$p(i)=\frac{e^{-\lambda}\lambda^i}{i!} \qquad i\geqslant 0$$

的随机变量 X 称为参数为 λ 的**泊松随机变量**. 如果进行很多次独立试验，每次成功的概率很小，则试验成功的次数可以近似地认为是泊松随机变量. 其期望和方差均等于参数 λ，即

$$E[X]=\mathrm{Var}(X)=\lambda$$

概率质量函数为

$$p(i)=p(1-p)^{i-1} \qquad i=1,2,\cdots$$

的随机变量 X 称为参数为 p 的**几何随机变量**. 这样的随机变量表示首次成功所需的试验次数，每次独立试验成功的概率为 p. 其期望和方差如下：

$$E[X]=\frac{1}{p} \qquad \mathrm{Var}(X)=\frac{1-p}{p^2}$$

概率质量函数为

$$p(i)=\binom{i-1}{r-1}p^r(1-p)^{i-r} \qquad i\geqslant r$$

的随机变量 X 称为参数为(r,p)的**负二项随机变量**. 这样的随机变量表示 n 次独立试验中第 r 次试验成功所需的试验次数，每次试验成功的概率为 p. 其期望和方差如下：

$$E[X]=\frac{r}{p} \qquad \mathrm{Var}(X)=\frac{r(1-p)}{p^2}$$

假设箱中有 N 个球，其中白球 m 个，现随机地从箱中取出 n 个球，所取白球个数可以认为是参数为(n,N,m)的**超几何随机变量**，其概率质量函数为

$$p(i)=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \qquad i=0,\cdots,m$$

若 $p=m/N$，其期望和方差如下：

$$E[X]=np \qquad \mathrm{Var}(X)=\frac{N-n}{N-1}np(1-p)$$

170

习题

1. 从装有 8 个白球、4 个黑球与 2 个黄球的箱中随机地取出 2 个球. 假定每取出 1 个黑球赢 2 美元，而取出 1 个白球输 1 美元. 以 X 表示赢的钱数，问 X 可能取哪些值? X 取这些值的概率各是多少?
2. 掷两个均匀的骰子，以 X 表示这两个骰子点数的乘积，试对 $i=1,2,\cdots$求出 $P\{X=i\}$.
3. 掷三个骰子，共有 $6^3=216$ 种可能结果，假定每一种结果是等可能的. 以 X 表示三个骰子的点数之和，试求 X 取各个可能值的概率.

4. 将 5 个男生与 5 个女生按照他们在一次考试中的成绩排队. 假设他们的成绩各不相等，且 10! 种排列是等可能的. 以 X 表示女生得到的最高名次(例如，当第一名是女生时 $X=1$，当第一名是男生而第二名是女生时 $X=2$)，试求 $P\{X=i\}, i=1,2,\cdots,10$.
5. 将一枚硬币抛 n 次，设 X 表示得到正面次数与反面次数之差，X 可能取哪些值?
6. 在上题中，假定此硬币是均匀的，对 $n=3$，问 X 取各个可能值的概率是多少?
7. 假定把一个骰子掷两次. 如下随机变量各取哪些可能值?
   (a) 两次掷出的最大点数.
   (b) 两次掷出的最小点数.
   (c) 两次掷出的点数之和.
   (d) 第一次掷出的点数减去第二次掷出的点数.
8. 若在上题中假定骰子是均匀的，对(a)～(d)分别求出随机变量取各个可能值的概率.
9. 试对有放回取球的情形重解例 1b.
10. 在例 1d 中，若已知我们赢了，试求赢得 i 美元的条件概率，并对 $i=1,2,3$ 求出概率值.
11. (a)从$\{1,2,\cdots,10^3\}$中随机选一个整数 N，每个数被选中的概率是相等的，试求 N 能被 3，5，7，15，105 这些数整除的概率. 若 k 足够大，用 10^k 代替 10^3，试问上述概率将怎样变化?
    (b)数论上的一个重要函数是默比乌斯(Möbius)函数 $\mu(n)$，其性质与数学上可能最重要的不可解问题——黎曼假设有关，对所有的正整数 n，该函数的定义如下：因子 n 可分解为几个素数. 如果有重复的素数，例如，$12=2\cdot2\cdot3$ 或 $49=7\cdot7$，则 $\mu(n)$定义为 0. 从$\{1,2,\cdots,10^k\}$中随机选一个整数 N，其中 k 很大. 当 $k\to\infty$时，试求 $P\{\mu(N)=0\}$.
    提示：为求 $P\{\mu(N)\neq0\}$，利用恒等式
$$\prod_{i=1}^{\infty}\frac{P_i^2-1}{P_i^2}=\left(\frac{3}{4}\right)\left(\frac{8}{9}\right)\left(\frac{24}{25}\right)\left(\frac{48}{49}\right)\cdots=\frac{6}{\pi^2}$$
    其中 P_i 是第 i 个最小的素数. (这里素数中不包括 1.)
12. 在"莫拉(Morra)二指"赌博中，两个赌徒各伸出 1 或 2 个手指，并同时猜对方伸出的手指数. 如果两人中只有一人猜对了，那么他赢得两人伸出的手指之和的钱数(单位：美元). 如果两人都猜对了或都没猜对，则谁也不赢谁的钱. 现考虑某赌徒，并设他在一局"莫拉二指"赌博中赢得的钱数为 X. [171]
    (a) 如果两赌徒的行为是相互独立的，而且设每一赌徒将伸出几个手指以及他猜测对方伸出几个手指总共 4 种结果是等可能的. 试问 X 取哪些可能值? 取这些值的相应概率各是多少?
    (b) 假定两赌徒的行为是相互独立的，而且每人猜对方伸几个手指就决定自己伸几个手指，且设每个赌徒伸出 1 或 2 个手指是等可能的. 试问 X 取哪些可能值? 取这些值的相应概率各是多少?
13. 某推销员以两种方案卖百科全书，第一种方案卖出的概率为 0.3，而第二种方案卖出的概率为 0.6. 每种方案等可能地卖精装版或普通版，其中精装版卖 1 000 美元，普通版卖 500 美元，设 X 表示卖出书的总收益，试求 X 的概率质量函数.
14. 5 个不同的数随机地分给编号为 1 至 5 的 5 个人，其中任意两个人比较他们的数，数大的获胜. 最初，编号为 1 和编号为 2 的两个人比较他们的数，获胜的一方再和编号为 3 的人比较，以此类推. 设 X 表示编号为 1 的人获胜的次数，试求 $P\{X=i\}$，$i=0$，1，2，3，4.
15. NBA 选秀包含 11 个队，这些队在当年的常规赛中成绩最差. 箱中共有 66 个球，每个球上写着球队的名称，其中 11 个球上写着最差球队名，10 个球上写着倒数第二球队名，9 个球上写着倒数第三球队名，以此类推. (1 个球上写着倒数第十一球队名.) 随机取出一个球，球上写的球队获得第一次选秀机会. 选出另外一个球，其球队名和第一次选秀的球队名不一样，该队获得第二次选秀机会. (如果取出的球写的球队名和第一次一样则扔掉，继续取球直到取出写不一样的球队名的球.) 最后，又取出

一个球，其球队名和前两次取出的球队名不一样，该队获得第三次选秀机会. 其余 4 到 11 次选秀机会属于没有获得前三次选秀机会的 8 支球队，并按他们在常规赛中成绩相反的顺序进行. 例如，如果最差的那支球队没有得到前三次选秀机会，则该球队会得到第 4 次选秀机会. 设 X 表示成绩最差的球队的选秀，试求 X 的概率质量函数.

16. 在上题中，设 1 号球上写着最差球队名，2 号球上写着倒数第二球队名，等等. 设 Y_i 表示得到第 i 次选秀机会的球队. 于是 $Y_1=3$ 表示选出的第 1 个球写着倒数第三球队名，试求下列概率质量函数：(a)Y_1；(b) Y_2；(c) Y_3.

17. 设 X 的分布函数为

$$F(b)=\begin{cases}0 & b<0\\ b/4 & 0\leqslant b<1\\ 1/2+(b-1)/4 & 1\leqslant b<2\\ 11/12 & 2\leqslant b<3\\ 1 & 3\leqslant b\end{cases}$$

(a) 求 $P\{X=i\}$，$i=1$，2，3. (b) 求 $P\left\{\frac{1}{2}<X<\frac{3}{2}\right\}$.

172

18. 将一枚均匀的硬币独立地抛 4 次，以 X 表示出现正面的次数，试画出随机变量 $X-2$ 的概率质量函数的图形.

19. 若 X 的分布函数为

$$F(b)=\begin{cases}0 & b<0\\ 1/2 & 0\leqslant b<1\\ 3/5 & 1\leqslant b<2\\ 4/5 & 2\leqslant b<3\\ 9/10 & 3\leqslant b<3.5\\ 1 & b\geqslant 3.5\end{cases}$$

试求 X 的概率质量函数.

20. 某赌博书上提到一种轮盘赌的“赢的策略”. 赌徒押红色 1 美元，若出现红的(概率为$\frac{18}{38}$)则赌徒赢 1 美元就离去. 若赌徒输了(发生的概率为$\frac{20}{38}$)，他在接下去的两次轮盘旋转中，每次额外再在红色下赌注 1 美元，然后离去. 设 X 表示赌徒赢了后离去.
(a) 试求 $P\{X>0\}$.
(b) 你确信这是一个赢的策略吗？试解释之.
(c) 试求 $E[X]$.

21. 有 4 辆公交车送 148 名学生从学校到体育场，每辆车分别有 40，33，25，50 名学生. 随机选出一名学生，设 X 表示这名学生所在的车上的学生数. 在 4 名公交车的司机中随机选出一名，设 Y 表示在这名司机的车上的学生数.
(a) 你认为 $E[X]$与 $E[Y]$哪个大？为什么？
(b) 试求 $E[X]$与 $E[Y]$.

22. 两个队进行一系列比赛，其中一队赢 i 场，比赛结束. 若甲队每场比赛获胜的概率为 p，且比赛结果是相互独立的，当(a) $i=2$；(b) $i=3$ 时，试求需要的比赛次数的数学期望. 证明在两种情形下，当 $p=\frac{1}{2}$时，这个值达到最大.

23. 某日用品目前每盎司[1]卖 2 美元，你有 1 000 美元，假定一周后这种日用品每盎司卖 1 美元或 4 美元、其中这两个概率是等可能的.

(a) 若要使这周末你能有最多的钱，你该选择哪种策略?

(b) 若要使这周末你能有最多的日用品数，你该选择哪种策略?

24. 甲、乙两人进行如下赌博：甲写下 1 或 2 中的某一个数让乙来猜，乙只能猜一个数. 若甲写的是 i 而乙猜对了，乙赢甲 i 元；若乙猜错了，则乙付给甲 3/4 元. 现假定乙的判断是随机的，他猜 1 的概率为 p，猜 2 的概率为 $1-p$. 试对(a)甲写的是 1 和(b)甲写的是 2 这两种情况分别求乙赢的钱数的期望值. p 取何值，可使乙的上述期望值的最小可能值达到最大? 这个最大值等于多少? (注意，乙赢的钱数的期望值不仅与 p 有关，而且与甲写的是哪个数有关.)

现考虑甲，假定甲写哪一个数也是随机的，他写 1 的概率为 q. 试对(c)乙猜的是 1 和(d)乙猜的是 2 这两种情况分别求出甲输的钱数的期望值.

173

q 取何值，可使甲的上述期望值的最大可能值达到最小? 试证这个最小值等于上面求出的乙赢的钱数的最小期望的最大值. 这个公共值叫做乙的博弈值. 上述结果称为极小极大定理，是博弈论这一数学分支中的基本结果，其一般形式由数学家冯·诺伊曼(John Von Neumann)首先证明. (注意，乙赢的钱数的期望值不仅与 p 有关，而且与甲写的是哪个数有关.)

25. 一种典型的"吃角子老虎"有三个转盘，每个转盘上印有 20 个标记(有樱桃，柠檬，李子，橘子，铃与棍棒). 典型的一套转盘如下表所示：

	转盘 1	转盘 2	转盘 3
樱桃	7	7	0
橘子	3	7	6
柠檬	3	0	4
李子	4	1	6
铃	2	2	3
棍棒	1	3	1
合计	20	20	20

根据这张表，转盘 1 上的 20 个标记是 7 个樱桃、3 个橘子，等等. 1 美元赌注的标准支付如下表所示：

转盘 1	转盘 2	转盘 3	支付
棍棒	棍棒	棍棒	60
铃	铃	铃	20
铃	铃	棍棒	18
李子	李子	李子	14
橘子	橘子	橘子	10
橘子	橘子	棍棒	8
樱桃	樱桃	任意	2
樱桃	非樱桃	任意	0
除此以外			−1

[1] 盎司的单位符号为 USfloz，1USfloz=29.573 53 cm³. ——编辑注

假设每个转盘独立地进行，求赌徒玩一次“吃角子老虎”所赢钱数的期望值.

26. 某个人从 1 到 10 随机地选取一个数，你试图用提问题的方式猜测他选取的数，而他只回答“是”或“不是”. 对下述两种情况，分别求出你需要提的问题数的期望值.
   (a) 你的第 i 个问题是“它是 i 吗?”($i=1, 2, 3, 4, 5, 6, 7, 8, 9, 10.$)
   (b) 你每提一个问题都试图排除当时所剩下的数的一半(或接近一半).
27. 某保险公司里一条保险规定，如果在一年内某事件 E 发生，该公司就要赔偿一笔钱 A. 该公司估计，在一年内 E 发生的概率为 p，那么为使他们收益的期望值为 A 的 10%，该公司应要求顾客交多少保险金?
28. 某盒内装有 20 件产品，其中 4 件次品，现随机地从盒中抽取 3 件产品，求在这 3 件产品中次品数的期望值.

174

29. 某机器损坏的原因有两种可能，为检查是否由第一种原因引起，需花费 C_1 美元. 如果查明是第一种原因，进一步排除故障的修理费为 R_1 美元. 同样地，对第二种原因需检查费 C_2 美元和修理费 R_2 美元. 设机器由于第一种及第二种原因损坏的概率分别为 p 及 $1-p$. 试问 p, C_i, $R_i(i=1, 2)$应满足什么条件，按照先查第一种原因后查第二种原因的次序来检修机器比倒过来的检修次序所需的费用的期望值小?
   提示：若一种检查失效，则必须进行另一种检查.
30. 某人抛一枚均匀硬币直到首次出现反面为止. 若第 n 次抛出反面，他赢得 2^n 美元. 设 X 表示其赢利，证明$E[X]=\infty$. 这个问题称为圣彼得堡(St. Petersburg)悖论.
   (a) 只赌一次，付 100 万元你愿意吗?
   (b) 如果允许你想赌几次就赌几次，直到你终止赌博为止，那么每赌一次付 100 万元你愿意吗?
31. 每晚不同的气象学家都会给出第二天有雨的概率. 为判断他们预测的好坏，按如下方法给他们打分：如果某气象学家说第二天有雨的概率为 p，则他会得到

$$1-(1-p)^2 \qquad \text{确实有雨}$$
$$1-p^2 \qquad \text{没有雨}$$

   在某时间跨度统计分数，认为得到最高分的气象学家是最好的预测者. 假定某气象学家知道这种方法，他想得到最高分. 如果此人确信明天下雨的概率为 p^*，试问 p 取什么值他能得到最高分?
32. 为了诊断在 100 人中是否有人患某种疾病，需要对他们验血. 方法是：分成每 10 人一组，将这 10 人的血液样本集中起来一块进行分析，而不是每人逐个地化验. 若化验结果为阴性，则 10 人只需进行一次化验即可. 若化验结果为阳性，就要对这 10 人再逐个地化验，该组总共要进行 11 次化验. 假定每人是否患此病是相互独立的，且每人患此病的概率都等于 0.1，试求对 100 个人所需化验的次数的期望值.（注意，这里我们假定集中起来的血液中，若至少有一人患此病，则化验结果呈阳性.）
33. 一报童卖报每份 15 美分，其成本为 10 美分. 但是，他不能把卖不出去的报纸退回. 如果他每日的销售量是以 $n=10$, $p=\frac{1}{3}$为参数的二项随机变量，则为使他的期望利润达到最大，他大体上应该购买多少份报纸?
34. 在例 4b 中，设百货商店因商品不对路，造成其成本每件增加 c 元.（通常称之为商誉费，即因顾客的需要得不到满足，造成商品失去信誉的损失费.）试求商店储备 s 件商品时的期望利润，并确定 s 使其期望利润达到最大.
35. 盒中有 5 红 5 蓝两种弹子，随机选出 2 个弹子，如果它们是同色的，你就会赢 1.1 美元；如果它们是不同色的，你会赢-1.1 美元(也就是说，输 1.1 美元). 试求
   (a) 你赢的期望值.

(b) 你赢的方差.

175

36. 在22题中，令 $i=2$. 试求比赛次数的方差，并证明当 $p=\frac{1}{2}$ 时，比赛次数最多.

37. 试求21题中 X 与 Y 的方差 $\mathrm{Var}(X)$ 及 $\mathrm{Var}(Y)$.

38. 若 $E[X]=1$，$\mathrm{Var}(X)=5$，试求

    (a) $E[(2+X)^2]$；(b) $\mathrm{Var}(4+3X)$.

39. 从装有3个白球与3个黑球的箱中取出一球，将它放回后再取一球，这样反复做下去. 取出的前4个球中恰有2个白球的概率是多少?

40. 一张考卷上有5道题，每道题列出3个可能答案，其中有一个答案是正确的. 某个学生靠猜测能答对至少4道题的概率是多少?

41. 某人自称有超感应力. 为对他考验，将一枚均匀的硬币抛10次，让他事先预测结果. 10次中他说对7次. 如果他没有超感应力，他将做出至少这样好的答案的概率是多少?

42. 假设飞行中飞机引擎损坏的概率为 $1-p$，且各引擎是否损坏是相互独立的. 如果要使一架飞机飞行必须有过半数的引擎正常工作，问对多大的 p 值，5个引擎的飞机比3个引擎的飞机更好?

43. 某通信系统传送数字0与1. 但由于静电干扰，传送的每个数字被错误接收的概率为0.2. 假定我们要传送一个由0与1组成的重要电报，为减少出错的机会，我们用00000代替0，而用11111代替1. 如果接收者用“过半”译码法，那么译出后的电报是错误的概率等于多少? 这里你做了怎样的独立性假设?

44. 某卫星系统由 n 个部件组成，某天至少 k 个部件运行则系统正常工作. 下雨天每个部件独立运行的概率为 p_1，而晴天此概率为 p_2. 若明天下雨的概率为 α，试问整个卫星系统正常工作的概率是多少?

45. 某个学生正准备参加一次重要的口试，且对口试那天他处于怎样的精神状态很关心. 因为他估计，如果他处于“最佳”状态，则主考人各自独立地给他及格的概率为0.8；若他处于“很坏”的状态，上述概率则降为0.4. 若假定有过半数主考人给他及格，他的口试才算及格，且这个学生已感到他处于“很坏”状态的可能性是处于“最佳”状态的两倍，问他是请求5位主考人还是3位主考人为好?

46. 假设某12个人组成的陪审团至少有9票才能判决被告有罪，并设陪审团对有罪的人投无罪票的概率为0.2，而对无罪的人投有罪票的概率为0.1. 如果各陪审员的行为是相互独立的，且65%的被告事实上有罪. 试求陪审团做出正确判决的概率. 被告中被判有罪的占百分之多少?

47. 某军事法庭任命9名法官. 但是，原告与被告的律师都有权拒绝任何法官出庭，被拒绝的法官立即离庭而且不再更换. 如果过半数法官投有罪票，则被告被判有罪，否则他被判无罪. 假设对一个事实上有罪的被告，各法官(独立地)投有罪票的概率为0.7，而对事实上无罪的被告，这个概率下降到0.3.

    (a) 当有(i)9，(ii)8，(iii)7名法官出庭时，一个有罪的被告被判有罪的概率各是多少?

176

    (b) 对一个无罪的被告重新做(a).

    (c) 假定原告律师不行使拒绝法官出庭的权利，并限定被告律师至多可拒绝两名法官出庭. 如果被告律师有60%的把握判断他的当事人事实上是有罪的，那么他拒绝几位法官为好?

48. 已知某公司生产的磁盘以概率0.01为次品，并设各磁盘是否为次品是相互独立的，这家公司将10张磁盘包成一包，并保证若发现某包内多于一个次品则可退款. 若有人买了3包，试问他恰好退回一包的概率是多少?

49. 某计算机硬件制造商生产的芯片有10%是次品，若我们订购100个此种芯片，试问订购的芯片中的次品数是二项随机变量吗?

50. 抛一枚偏重的硬币出现正面的概率为 p，抛了10次，其中6次出现正面，试求条件概率，如果前三次的结果是

    (a) H，T，T(表示第一次是正面，第二次与第三次均出现反面).

(b) T, H, T.

51. 某杂志每页出现印刷错误的期望值为 0.2，试求你所读的下页书中包含(a) 0，(b) 2 处或更多处出现印刷错误的概率. 解释你的推理.
52. 每月全世界飞机失事的平均数为 3.5，试求下列概率：
    (a) 下个月至少两架飞机失事.
    (b) 下个月最多一架飞机失事.
    解释你的推理.
53. 去年纽约大约有 80 000 人结为夫妇. 估计下列概率：
    (a) 其中至少两对人生于 4 月 30 日.
    (b) 其中至少两对人同一天过生日.
    并说明原因.
54. 某高速公路每周废弃的小汽车平均为 2.2 辆，试求下列概率：
    (a) 下周没有小汽车废弃.
    (b) 下周至少有两辆小汽车废弃.
55. 某打印处有 2 名打字员，第一个打字员每篇文章平均出错 3 处，而第二个打字员每篇文章平均出错 4.2 处. 如果两个打字员等可能地打你的文章，试求没有错误的近似概率.
56. 试问需要多少人，他们中的任何一人和你生日相同的概率大于$\frac{1}{2}$?
57. 设在某高速公路上每天发生事故的次数是以 $\lambda=3$ 为参数的泊松随机变量.
    (a) 试求今天出了 3 次或更多次事故的概率.
    (b) 假定今天至少出了一次事故，在此条件下重做(a).
58. 对于下列情形，试求二项随机变量的概率及其泊松近似，并进行比较.
    (a) $n=8$，$p=0.1$ 时的 $P\{X=2\}$.
    (b) $n=10$，$p=0.95$ 时的 $P\{X=9\}$.
    (c) $n=10$，$p=0.1$ 时的 $P\{X=0\}$.
    (d) $n=9$，$p=0.2$ 时的 $P\{X=4\}$.
59. 如果你买了 50 张彩票，每张中彩的机会是$\frac{1}{100}$，问你将有(a)至少一张，(b)正好一张，(c)至少两张彩票中彩的(近似)概率是多少?

[177]

60. 一个人在一年内患感冒的次数是以 $\lambda=5$ 为参数的泊松随机变量. 假定正在销售的一种新特效药(含大量维生素 C)对 75%的人来说可将上述泊松参数减小到 $\lambda=3$，而对另外 25%的人来说则是无效的. 设某人使用此药一年，在试用期患了两次感冒，问此药对他有效的可能性是多少?
61. 随机地分发纸牌时，被发到一手满堂红(三张同点加一对)的概率近似地为 0.001 4. 试求将要发给你的1 000手牌中至少有 2 个满堂红的近似概率值.
62. n 对夫妇随机地围坐在一张圆桌旁，试求没有一个妻子坐在她丈夫旁边的近似概率. 计算 $n=10$ 时的概率，并与第 2 章例 5n 作比较.
63. 人们以每 2 分钟 1 人的速度进入某赌场. 试问
    (a) 在 12：00 至 12：05 之间没有人进入赌场的概率是多少?
    (b) 在这段时间内至少有 4 人进入赌场的概率是多少?
64. 在某个州，自杀率是每月每 10 万人中有一人自杀.
    (a) 试求在这个州拥有 40 万人口的某城市中，在某月内自杀者超过 7 人的概率.
    (b) 在一年期间，至少有两个月这个城市的自杀者超过 7 人的概率是多少?
    (c) 把现在这个月编为 1 号，试问首次出现超过 7 人自杀的月份是编号为 $i(i\geqslant1)$的月份的概率是多

少?

你做了怎样的假设?

65. 某部队有 500 名士兵，每人独立地患一种病的概率为 $1/10^3$. 确诊此病需进行一项血液化验，为方便起见，直接把 500 人的血液混合在一起化验.
    (a) 试问化验结果是阳性(也就是说至少一人有这种病)的(近似)概率为多少?
    (b) 在这种条件下，试问多于一个人患这种病的概率是多少? 500 人中有一人是琼斯，他知道他患这种疾病.
    (c) 琼斯认为多于一个人患这种病的概率是多少? 若化验结果是阳性，这时医生决定对每人的血液分别进行化验. 前 $i-1$ 次化验结果是阴性的，第 i 个人，(琼斯)是阳性的.
    (d) 作为 i 的一个函数，试问其余人任何一人患这种病的概率是多少?
66. 考虑由 38 个数字(1 至 36，0 及双 0)组成轮盘. 如果史密斯总是把赌注下在 1 到 12 这些数的某一个上，问他将在前五局中全输掉的概率是多少? 他在第四局首次赢钱的概率是多少?
67. 两个运动队进行一系列比赛，约定先胜四场的队为全胜. 假定两队中一队比另一队强，每场比赛强队获胜的概率为 0.6，且各场比赛的结果是相互独立的. 试求强队经 $i(i=4,5,6,7)$场比赛获全胜的概率. 并比较按上述规则与按三战两胜规则强队获全胜的概率大小.

178

68. 在上题中，若两队实力相当，每队赢的概率均为$\frac{1}{2}$，试求需要的比赛场次的期望值.
69. 某记者收到一份他可以采访的人名单. 如果记者需采访 5 个人，而名单上每个人(独立地)接受采访的概率均为 2/3. 如果名单上有(a)5 个人，(b)8 个人，记者能得到他需要的采访人数的概率是多少? 对于(b)，记者能与名单上的正好(c)6 个人，(d)7 个人交谈的概率是多少?
70. 连续抛一枚均匀的硬币，直到正面出现 10 次为止，令 X 为其中反面出现的次数，试求 X 的概率质量函数.
71. 如果左、右衣袋里的火柴盒最初各有 N_1，N_2 根，试解巴拿赫火柴问题(例 8e).
72. 在巴拿赫火柴问题(例 8e)中，试求当第一盒用完的时候(不是发现它是空盒时)另一盒恰有 k 根火柴的概率.
73. 箱中有 4 个白球与 4 个黑球. 现随机取出 4 个球，若其中黑、白球各 2 个则停止；否则把这 4 个球放回后再随机取 4 个. 这样继续下去，直到取出的球中恰有 2 个白球为止. 问正好取 n 次的概率是多少?
74. 假定 100 件产品中有 6 件次品、94 件正品. 令 X 表示随机地抽取的 10 件产品中的次品数，试求 (a) $P\{X=0\}$，(b) $P\{X>2\}$.
75. 流行于内华达州赌场的一种赌博叫“基诺”(Keno)，其赌法如下：赌场从 1 至 80 号中随机地取出 20 个号码，一个赌徒从这 80 个号码中任取 1 至 15 个号码. 如果赌徒取出的号码中部分与赌场所取出的 20 个号码相同，则他获胜. 赢得的钱数是赌徒所取号码的个数与其中配上对的号码个数的函数. 例如，赌徒只取一个号码时，若此号码在赌场所取的 20 个号码之中，则他每下 1 美元赌注赢 2.2 美元. (由于此时赌徒的获胜概率是 1/4，所以公平的赢钱数显然应是每 1 美元赌注赢 3 美元.)当赌徒取出 2 个号码时，如果这 2 个号码都在赌场所取的 20 个号码之中，则每下 1 美元赌注(不公平地)赢 12 美元.
    (a) 此时公平的赢钱数应是多少?

    令 $P_{n,k}$表示赌徒取出的 n 个号码之中恰有 k 个在赌场取出的 20 个号码之中的概率.
    (b) 求出 $P_{n,k}$.
    (c) 在基诺赌博中，最典型的赌法是赌徒取 10 个号码. 这种情况下赌场付出的钱数如下表，试求出此表的最后一栏.

赌 10 个号码的基诺赌博的赢钱数

配对个数	每 1 美元赌注的赢钱数	公平的赢钱数
0～4	−1	
5	1	
6	17	
7	179	
8	1 299	
9	2 599	
10	24 999	

179

76. 在例 8i 中，一个含有 $i(i=1, 4)$ 个次元件的包，被采购员拒绝的百分数是多少？已知某包被拒绝，其中含有 4 个次元件的条件概率是多少？
77. 某采购员购买 20 个一包的晶体管，他的方法是：从每包中随机地选择 4 个，只有当 4 个全是正品时才买下这一包，设各晶体管是次品的概率独立地为 0.1，问被拒绝的包数占多大比例？

理论练习

1. 有 N 种不同类型的票券，每次只能得到一张，且各次的选择是相互独立的，以概率 P_i 得到第 $i(i=1, \cdots, N)$ 个类型. 设 T 表示需要通过选择得到每种类型至少一张的数目. 计算 $P\{T=n\}$.
   提示：用类似例 1e 的方法求解.
2. 运用概率的连续性计算 $\lim_{b\to-\infty} P\{X\leqslant b\}$.
3. 试用 X 的分布函数表示 $P\{X\geqslant a\}$.
4. 若 X 有分布函数 F，问 e^X 的分布函数是什么？
5. 若 X 有分布函数 F，问随机变量 $\alpha X+\beta$ 的分布函数是什么？其中 α 与 β 是常数，$\alpha\neq 0$.
6. 对任一非负整数值随机变量 N，证明

$$E[N] = \sum_{i=1}^{\infty} P\{N \geqslant i\}$$

   提示：$\sum_{i=1}^{\infty} P\{N\geqslant i\} = \sum_{i=1}^{\infty}\sum_{k=i}^{\infty} P\{N=k\}$ 交换和的次序即得.
7. 对任一非负整数值随机变量 N，证明

$$\sum_{i=0}^{\infty} iP\{N>i\} = \frac{1}{2}(E[N^2]-E[N])$$

   提示：$\sum_{i=0}^{\infty} iP\{N>i\} = \sum_{i=0}^{\infty} i \sum_{k=i+1}^{\infty} P\{N=k\}$ 交换和的次序即得.
8. 设 X 满足

$$P\{X=1\}=p=1-P\{X=-1\}$$

   求 $c\neq 1$，使得 $E[c^X]=1$.
9. 设随机变量 X 的数学期望为 μ，方差为 σ^2. 计算

$$Y=\frac{X-\mu}{\sigma}$$

   的数学期望与方差.
10. 设 X 是以 (n, p) 为参数的二项随机变量，证明

180

$$E\left[\frac{1}{X+1}\right]=\frac{1-(1-p)^{n+1}}{(n+1)p}$$

11. 考虑 n 个独立的连续试验，成功的概率为 p，如果有 k 次成功，试证明 k 次成功、$n-k$ 次失败共有 $n!/[k!(n-k)!]$种可能排列.
12. 共有 n 个元件线性排列，每个元件以概率 p 独立运行，试问没有两个相邻元件都不运行的概率是多少?

    提示：以有缺陷的元件数为条件，用第 1 章例 4c 的结论来解.
13. 设 X 是以(n,p)为参数的二项随机变量，怎样的 p 值能使 $P\{X=k\}(k=0,1,\cdots,n)$取到最大? 这是一个当以(n,p)为参数的随机变量 X 的观测值为k 时，利用统计方法去估计 p 的例子. 如果假定 n 是已知的，我们选取使 $P\{X=k\}$达到最大的 p 作为 p 的估计值. 这种方法称为最大似然估计.
14. 一个家庭有 n 个子女的概率为 αp^n，其中 $n\geqslant 1,\alpha\leqslant\frac{1-p}{p}$. 试问：

    (a) 没有孩子的家庭占多大比例?

    (b) 如果每个孩子为男孩或女孩是等可能的(且相互独立)，那么有 k 个男孩(不管有几个女孩)的家庭占多大比例?
15. 一枚硬币抛出正面的概率为 p，先将它独立地抛 n 次. 试证出现偶数个正面的概率为$\frac{1}{2}[1+(q-p)^n]$，其中 $q=1-p$. 可先证明等式

$$\sum_{i=0}^{[n/2]}\binom{n}{2i}p^{2i}q^{n-2i}=\frac{1}{2}[(p+q)^n+(q-p)^n]$$

    其中$[n/2]$是不大于 $n/2$ 的最大整数，然后利用此等式来证明结论. 试将这个习题与第 3 章理论练习 15 作一比较.
16. 设 X 是参数为λ 的泊松随机变量，试证当 i 增大时，$P\{X=i\}$先是单调递增然后单调递减，并当 i 为不超过λ 的最大整数时取到最大值.

    提示：考虑 $P\{X=i\}/P\{X=i-1\}$.
17. 设 X 是参数为λ 的泊松随机变量.

    (a) 证明

$$P\{X\text{ 是偶数}\}=\frac{1}{2}[1+e^{-2\lambda}]$$

    利用理论练习 15 以及泊松随机变量与二项随机变量的关系来证明.

    (b) 运用 $e^{-\lambda}+e^{\lambda}$的展开式直接验证上述结论.
18. 设 X 是参数为λ 的泊松随机变量，试问对任意 $k\geqslant 0$，λ 取何值使得 $P\{X=k\}$达到最大值?
19. 设 X 是参数为λ 的泊松随机变量，试证

$$E[X^n]=\lambda E[(X+1)^{n-1}]$$

    利用此结论计算 $E[X^3]$.
20. 考虑n 枚硬币，每次独立地抛出正面的概率为 p. 若 n 很大，p 很小，令 $\lambda=np$. 假定抛 n 枚硬币，至少出现 1 次正面试验结束，若没有则继续试验. 也就是说，n 枚硬币至少一枚出现正面第一次试验停止. 设 X 表示正面出现的次数. 下列求 $P\{X=1\}$的方法哪一个是正确的? 在所有情形下，设Y 是参数为λ 的泊松随机变量.

181

    (a) 因为抛 n 枚硬币出现正面的总数近似为参数为λ 的泊松随机变量，所以有

$$P\{X=1\}\approx P\{Y=1\}=\lambda e^{-\lambda}$$

    (b) 因为抛 n 枚硬币出现正面的总数近似为参数为λ 的泊松随机变量，而且只有此总数是正的我们才停止试验，所以有

$$P\{X=1\}\approx P\{Y=1\mid Y>0\}=\frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}$$

(c) 因为至少一枚硬币出现正面，所以若其余 $n-1$ 枚硬币再没有一枚出现正面则 $X=1$. 因为 $n-1$ 枚硬币出现正面的次数近似为参数为 $(n-1)p\approx\lambda$ 的泊松随机变量，所以有

$$P\{X=1\}\approx P\{Y=0\}=e^{-\lambda}$$

21. 令 E_{ij} 表示第 i 人与第 j 人生日相同的事件. 假定随机选出 n 个人，每个人都等可能地在一年 365 天中的任何一天过生日. 试求：

(a) $P(E_{3,4}|E_{1,2})$；(b) $P(E_{1,3}|E_{1,2})$；(c) $P(E_{2,3}|E_{1,2}\cap E_{1,3})$.

22. 箱中有 $2n$ 个球，其中两个编号为 1，两个编号为 2，…，两个编号为 n. 现无放回地每次连续取出两个球. 设 T 表示一次取出的两个球编号相同.（若每次没有取出编号相同的球，则 T 等于无穷大）. 对 $0<\alpha<1$，试证

$$\lim_n P\{T>\alpha n\}=e^{-\alpha/2}$$

为证明此结论，令 M_k 表示前 k 次取出编号相同的球的次数，其中 $k=1,\cdots,n$.

(a) 讨论当 n 很大时，M_k 可(近似)看作是 k 次独立试验成功的次数.

(b) 当 n 很大时，试求 $P\{M_k=0\}$.

(c) 根据变量 M_k 的一个值写出 $\{T>\alpha n\}$.

(d) 给出 $P\{T>\alpha n\}$，试验证极限概率.

23. 假定某时期事件发生的次数是参数为 λ 的泊松随机变量. 如果每一个发生了的事件被计数的概率为 p，而且事件是否被计数是相互独立的，试证被计数的事件个数是参数为 λp 的泊松随机变量. 另外，给出这一结果的直观解释.

作为上述结果的一个应用，假定某个给定区域内铀矿的个数是参数为 $\lambda=10$ 的泊松随机变量. 如果在某固定时期，独立地找到每一个矿的概率为 1/50，试求在这段时间内下列事件发生的概率：(a)正好找到一个矿，(b)至少找到一个矿，(c)至多找到一个矿.

24. 试证

$$\sum_{i=0}^{n}e^{-\lambda}\frac{\lambda^i}{i!}=\frac{1}{n!}\int_{\lambda}^{\infty}e^{-x}x^n\,dx$$

提示：运用分部积分法.

182

25. 设 X 为几何随机变量，试用解析方法证明

$$P\{X=n+k|X>n\}=P\{X=k\}$$

利用几何随机变量的含义对上式为什么成立给出一个口头说明.

26. 设 X 是参数为 (r,p) 的负二项随机变量，Y 是参数为 (n,p) 的二项随机变量，试证

$$P\{X>n\}=P\{Y<r\}$$

提示：可用解析方法证明，上述结论等价于

$$\sum_{i=n+1}^{\infty}\binom{i-1}{r-1}p^r(1-p)^{i-r}=\sum_{i=0}^{r-1}\binom{n}{i}p^i(1-p)^{n-i}$$

也可利用随机变量的概率解释来证明. 也就是说，后一种情形是考虑试验成功的概率为 p 的独立试验事件序列，运用这些事件的结果表示 $\{X>n\}$ 与 $\{Y<r\}$.

27. 对一个超几何随机变量，试求

$$P\{X=k+1\}/P\{X=k\}$$

28. 编号为 1 至 N 的球放在一个箱中，假定从中无放回地随机取出 $n(n\leqslant N)$ 个，以 Y 表示被取出的球的最大编号.

(a) 求 Y 的概率质量函数.

(b) 导出 $E[Y]$ 的表达式，并利用费马组合恒等式(第 1 章理论练习 11)进行简化.

29. 罐中有编号为 $1,2,\cdots,n+m$ 的筹码. 先从中取出 n 个，在被取出的筹码中，以 X 表示其编号比所有

末取出的筹码编号都大的筹码个数，试求 X 的概率质量函数.

30. 罐中有 n 个筹码，假定某人从罐中不断地取出一个筹码，每次取出一个后放回，再取下一个. 这样连续进行直到他取出先前已被取出过的一个筹码为止. 以 X 表示他取筹码的次数，试求 X 的概率质量函数.
31. 利用式(8-5)证明式(8-6).
32. 从一个 n 个元素的集合中随机地选择一个子集，并假定所有子集等可能地被选择. 令 X 表示在所选择的子集中元素的个数，使用第 1 章理论练习 12 给出的恒等式证明

$$E[X]=\frac{n}{2-\left(\frac{1}{2}\right)^{n-1}}$$

$$\mathrm{Var}(X)=\frac{n\cdot 2^{2n-2}-n(n+1)2^{n-2}}{(2^n-1)^2}$$

证明当 n 很大时，

$$\mathrm{Var}(X)\sim\frac{n}{4}$$

183

当 $n\to\infty$ 时，上述比率趋近于 1. 试将它与 $P\{Y=i\}=1/n(i=1,\cdots,n)$ 时 $\mathrm{Var}(Y)$ 的极限形式进行比较.

33. 箱中最初有一个红球和一个蓝球，随机取出一球，然后再加上一个颜色相同的球一起放回箱中. 设 X 表示第一次取出蓝球的次数. 例如，如果第一次取出红球，第二次取出蓝球，则 $X=2$.
    (a) 计算 $P\{X>i\}$，$i\geqslant 1$.
    (b) 试证蓝球以概率 1 最终将会被取出.（即证明 $P\{X<\infty\}=1$.）
    (c) 计算 $E[X]$.

自测题与练习

1. 假定随机变量 X 表示某棒球运动员接下来三次击球击中的次数，若 $P\{X=1\}=0.3$，$P\{X=2\}=0.2$，且 $P\{X=0\}=3P\{X=3\}$，试求 $E[X]$.
2. 假定 X 取值 0，1，2. 对某个常数 c，$P\{X=i\}=cP\{X=i-1\}$，$i=1,2$，试求 $E[X]$.
3. 抛一枚硬币直到出现两次正面或反面停止，出现正面的概率为 p，试求抛硬币次数的期望值.
4. 某社团由 m 个家庭组成，其中第 n_i 个家庭有 i 个孩子，$\sum_{i=1}^{r}n_i=m$. 若随机选出一个家庭，令 X 表示选中的那个家庭的孩子数. 若从 $\sum_{i=1}^{r}in_i$ 个孩子中随机选出一个孩子，令 Y 表示选中的那个孩子所在的家庭的孩子数. 试证 $E[Y]\geqslant E[X]$.
5. 假定 $P\{X=0\}=1-P\{X=1\}$，若 $E[X]=3\,\mathrm{Var}(X)$，试求 $P\{X=0\}$.
6. 盒中有甲、乙两枚硬币. 其中甲硬币出现正面的概率为 0.6，乙硬币出现正面的概率为 0.3. 随机选出一枚硬币，然后抛该枚硬币. 不知道选出的是哪一枚硬币，若出现正面你就可以赢 10 美元，反之输 10 美元. 假定某知情人会卖给你一些信息(价值为 C)，该信息是你选出的是哪一枚硬币. 若你买了此信息，则你的期望收益是多少? 若你买了此信息，赌注为 x 美元，最后你可能得到 $x-C$ 或 $-x-C$ 美元（也就是说，后者表示输了 $x+C$ 美元)，问 C 取多大值，值得买此信息?
7. 某慈善家在红纸上写了一个正数 x，出示给公证人，然后把红纸反过来放到桌上. 公证人抛一枚均匀的硬币，若出现正面，他在蓝纸上写下 $2x$，若出现反面，则写下 $x/2$，然后把蓝纸反过来放在桌上. 事先不知道 x 值或抛硬币的结果，你可以选择翻阅红纸或蓝纸. 选好以后，观察纸上的数目，你可能获得奖金，可能是这张纸上的数目，也可能是另一张纸上的数目. 例如，如果你选择翻阅蓝纸，纸上的数目是

100，则你可能得到 100 元，或者会得到红纸上的数目(200 或 50 元)．假定你希望得到的奖金数目很大.

184

(a) 讨论你先选择翻阅红纸的情况，因为如果这样做，则不管观察到什么数，总是得到蓝纸上的数目.

(b) 若 y 是一个非负数，考虑下述策略．选择翻阅蓝纸，若这个值至少是 y，则会得到这个数目．若小于 y，则会得到红纸上的数目．设 $R_y(x)$表示慈善家写下 x 的值，当你采用这个策略时所得到的奖金值．试求 $E[R_y(x)]$．注意 $E[R_0(x)]$表示慈善家写下 x 的值，当你采用这个策略时总是选择翻阅蓝纸得到的数目.

8. $B(n, p)$表示参数为 n，p 的二项随机变量，试证

$$P\{B(n, p)\leqslant i\}=1-P\{B(n, 1-p)\leqslant n-i-1\}$$

提示：成功的次数小于等于 i 等价于失败次数的什么表示?

9. 设 X 是数学期望为 6、方差为 2.4 的二项随机变量，试求 $P\{X=5\}$.

10. 编号为 1 到 n 的球放在一个箱中，无放回地随机依次取出 m 个球，试求 $P\{X=k\}$，其中 $k=1, 2, \cdots, m$，X 表示被取出球的最大编号.

提示：先求 $P\{X\leqslant k\}$.

11. 甲、乙两个队进行一系列比赛，先胜 3 场的队获胜，假定每场甲队获胜的概率为 p，且各场比赛结果是相互独立的．试求下列概率：

(a) 已知甲队第 1 场获胜，求最后获胜的条件概率.

(b) 已知甲队最后获胜，求第 1 场获胜的条件概率.

12. 当地一个足球队会进行 5 场比赛．如果这个周末比赛赢了，则会参加联赛高级赛的最后 4 场比赛；如果输了，则会参加低级赛的最后 4 场比赛．在高级赛中该队每场赢的概率为 0.4，在低级赛中该队每场赢的概率为 0.7．该队在这个周末赢得比赛的概率为 0.5，在最后 4 场比赛赢得 3 场的概率是多少?

13. 某地区在一年内平均刮 5.2 次狂风，试问在这一年内有 3 次或更少次狂风的概率是多少?

14. 鸡下的蛋数是参数为 λ 的泊松随机变量，其中鸡蛋下在树叶上，树叶上有一种某类型的昆虫，只有这个变量是正数时才能观察到，这是因为如果这个变量等于 0，我们不知道昆虫在树叶上．若令 Y 表示观察到的鸡蛋数，则

$$P\{Y=i\}=P\{X=i\mid X>0\}$$

其中 X 是参数为 λ 的泊松随机变量，试求 $E[Y]$.

15. 某轮盘赌博一直继续到他赢了 4 次赌博为止，此时他赢 5 美元.

(a) 他玩 9 次的概率是多少?

(b) 若他赢了 4 次赌博，其期望收益是多少?

提示：每次他以概率$\frac{18}{38}$赢 5 美元，以概率$\frac{20}{38}$输 5 美元.

16. 三人去喝咖啡，他们通过抛硬币来决定谁来买单，如果某个人得到了与其余两人不同的结果，则他就是“奇数人”来买单．如果没出现“奇数人”，就继续抛，直到出现“奇数人”为止．试求下列概率：

185

(a) 恰好抛 3 次.

(b) 多于 4 次.

17. 若 X 是参数为 p 的几何随机变量，试证

$$E[1/X]=\frac{-p\log p}{1-p}$$

提示：可先求 $\sum_{i=1}^{\infty}a^i/i$ 的表达式，为求这个表达式，可写成 $a^i/i=\int_0^a x^{i-1}\mathrm{d}x$，然后交换和的次序并用分部积分即得.

18. 假定

$$P\{X=a\}=p \qquad P\{X=b\}=1-p$$

(a) 试证$\dfrac{X-b}{a-b}$是伯努利随机变量.

(b) 试求 $\mathrm{Var}(X)$.

19. 每局比赛赢的概率为 p，如果你计划玩 5 局，但是若你第 5 局赢了则继续玩直到你有 1 局输为止.

(a) 试求你玩的期望局数.

(b) 试求你输的期望局数.

20. 箱中最初有 N 个白球和 M 个黑球，每次无放回地随机取出球，试求在取出 m 个黑球之前取出 n 个白球的概率，其中 $n\leqslant N$，$m\leqslant M$.

186

第 5 章 连续型随机变量

5.1 引言

在第 4 章中，我们讨论了离散型随机变量，即随机变量的可能取值是有限个或可数无穷多个. 但是，还有一些随机变量，它们的可能取值是不可数的. 例如，火车到达某车站的时间与晶体的寿命. 用 X 表示这样的随机变量，如果存在一个定义在 $x\in(-\infty,+\infty)$ 上的非负函数 f，使得对任意的实数集 B[㊀] 有

$$P\{X\in B\}=\int_B f(x)\mathrm{d}x \tag{1-1}$$

我们就称 X 是连续[㊁]型随机变量，函数 $f(x)$ 叫做随机变量 X 的概率密度函数(probability density function)(见图 5-1).

换言之，式(1-1)说明 X 在 B 中的概率可通过概率密度函数在集合 B 中积分获得. 由于 X 必须取某些值，故 f 必须满足

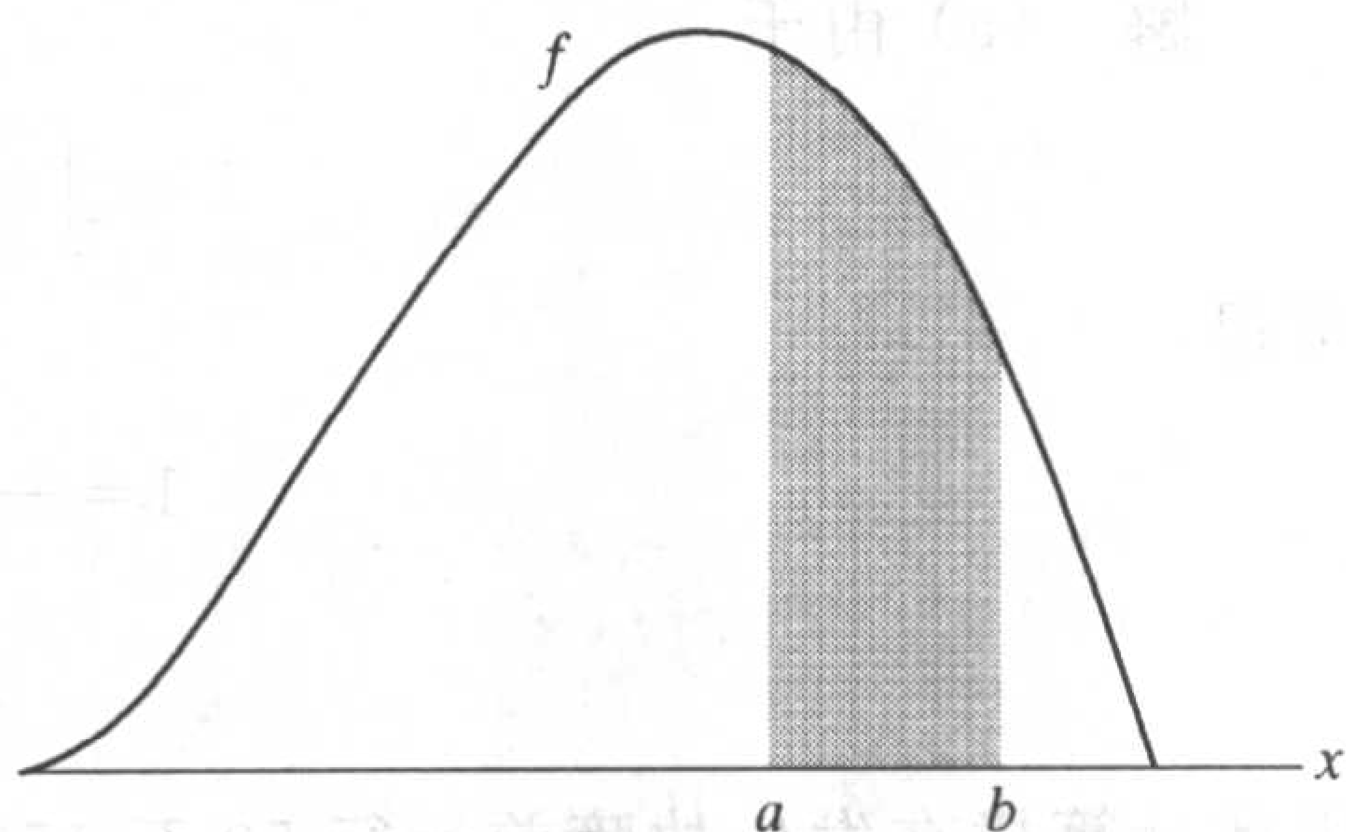
$$1=P\{X\in(-\infty,+\infty)\}=\int_{-\infty}^{+\infty}f(x)\mathrm{d}x$$

所有关于 X 的概率陈述都可以用 f 来表达. 例如，令 $B=[a,b]$，由式(1-1)可得

$$P\{a\leqslant X\leqslant b\}=\int_a^b f(x)\mathrm{d}x \tag{1-2}$$

如果在式(1-2)中令 $a=b$，则有

$$P\{X=a\}=\int_a^a f(x)\mathrm{d}x=0$$



$P(a\leqslant X\leqslant b)$ = 阴影部分的面积

图 5-1 概率密度函数

换句话说，这个式子说明连续型随机变量取某个固定值的概率为 0. 因此，对连续型随机变量，

$$P\{X<a\}=P\{X\leqslant a\}=F(a)=\int_{-\infty}^a f(x)\mathrm{d}x$$

例 1a 设 X 是连续型随机变量，其概率密度函数为

$$f(x)=\begin{cases}C(4x-2x^2) & 0<x<2\\ 0 & \text{其他}\end{cases}$$

(a) 求 C 的值.

(b) 求 $P\{X>1\}$.

解 (a) 由于 f 是概率密度函数，因此满足 $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$，从而

$$C\int_0^2(4x-2x^2)\mathrm{d}x=1$$

㊀ 实际上，由于技术上的原因，式(1-1)只有当 B 是可测集时成立，幸运的是实际中感兴趣的集合都包括在内了.

㊁ 有时称为绝对连续.

$$C\left[2x^2-\frac{2x^3}{3}\right]\Big|_{x=0}^{x=2}=1$$

$$C=\frac{3}{8}$$

187~188 因此

(b) $P\{X>1\}=\int_1^{+\infty}f(x)\mathrm{d}x=\frac{3}{8}\int_1^2(4x-2x^2)\mathrm{d}x=\frac{1}{2}$. ■

例 1b 某计算机在发生故障前运行的总时间(单位：小时)是一个连续型随机变量，其密度函数为

$$f(x)=\begin{cases}\lambda e^{-\frac{x}{100}} & x\geqslant 0\\ 0 & x<0\end{cases}$$

求下列事件发生的概率：

(a) 这个计算机在发生故障前能运行 50 至 150 小时.

(b) 它的运行时间少于 100 小时.

解 (a) 由于

$$1=\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=\lambda\int_0^{\infty}e^{-\frac{x}{100}}\mathrm{d}x$$

可得

$$1=-\lambda(100)e^{-\frac{x}{100}}\Big|_0^{\infty}=100\lambda$$

$$\lambda=\frac{1}{100}$$

因此计算机在发生故障前运行 50 至 150 小时的概率为

$$P\{50<X<150\}=\int_{50}^{150}\frac{1}{100}e^{-\frac{x}{100}}\mathrm{d}x=-e^{-\frac{x}{100}}\Big|_{50}^{150}=e^{-\frac{1}{2}}-e^{-\frac{3}{2}}\approx 0.384$$

(b) 类似可得

$$P\{X<100\}=\int_0^{100}\frac{1}{100}e^{-\frac{x}{100}}\mathrm{d}x=-e^{-\frac{x}{100}}\Big|_0^{100}=1-e^{-1}\approx 0.633$$

换句话说，计算机有 63.3%的可能会在工作不到 100 小时的情况下死机. ■

例 1c 某型号收音机电子管的寿命(单位:小时)是一个随机变量，其概率密度函数为

$$f(x)=\begin{cases}0 & x\leqslant 100\\ \frac{100}{x^2} & x>100\end{cases}$$

装有 5 个电子管的收音机在使用的前 150 小时之内恰有 2 个管子需要更换的概率是多少？

189 假定表示第 i 个电子管在这段时间内需要更换的事件 $E_i(i=1,2,3,4,5)$是独立的.

解 因为

$$P(E_i)=\int_0^{150}f(x)\mathrm{d}x=100\int_{100}^{150}x^{-2}\mathrm{d}x=\frac{1}{3}$$

故由事件 E_i 的独立性可得所求的概率为

$$\binom{5}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3=\frac{80}{243}$$ ■

累积分布函数 F 与概率密度函数 f 之间的关系可表示为

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)\mathrm{d}x$$

对上式两边求导得到

$$\frac{\mathrm{d}}{\mathrm{d}a}F(a) = f(a)$$

即密度函数是累积分布函数的导数. 可由式(1-2)给出密度函数如下更为直观的解释：当 ε 很小且 $f(\cdot)$ 在 $x=a$ 处连续时，

$$P\left\{a - \frac{\varepsilon}{2} \leqslant X \leqslant a + \frac{\varepsilon}{2}\right\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)\mathrm{d}x \approx \varepsilon f(a)$$

换句话说，X 将取值于以 a 点为中心、以 ε 为长度的小区间内的概率近似等于 $\varepsilon f(a)$. 因此我们看到 $f(a)$ 是这个随机变量以多大可能性落在 a 附近的一个度量.

5.2 连续型随机变量的数学期望与方差

在第 4 章中定义了离散型随机变量 X 的期望值为

$$E[X] = \sum_{x} xP\{X = x\}$$

设 X 是连续型随机变量，其概率密度函数为 $f(x)$，由于当 $\mathrm{d}x$ 很小时

$$f(x)\mathrm{d}x \approx P\{x \leqslant X \leqslant x + \mathrm{d}x\}$$

190

显而易见，可类似定义 X 的期望值为

$$E[X] = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$$

例 2a 设 X 的概率密度函数为

$$f(x) = \begin{cases} 2x & \text{若 } 0 \leqslant x \leqslant 1 \\ 0 & \text{其他} \end{cases}$$

求 $E[X]$.

解

$$E[X] = \int xf(x)\mathrm{d}x = \int_{0}^{1} 2x^2\mathrm{d}x = \frac{2}{3}$$

■

例 2b 设 X 的概率密度概率函数为

$$f(x) = \begin{cases} 1 & \text{若 } 0 \leqslant x \leqslant 1 \\ 0 & \text{其他} \end{cases}$$

求 $E[\mathrm{e}^X]$.

解 令 $Y = \mathrm{e}^X$. 先确定 Y 的分布函数 F_Y. 现在对 $1 \leqslant x \leqslant \mathrm{e}$，有

$$\begin{aligned} F_Y(x) &= P\{Y \leqslant x\} = P\{\mathrm{e}^X \leqslant x\} \\ &= P\{X \leqslant \log(x)\} = \int_{0}^{\log(x)} f(y)\mathrm{d}y = \log(x) \end{aligned}$$

对 $F_Y(x)$ 求导，得到 Y 的概率密度函数为

$$f_Y(x) = \frac{1}{x} \qquad 1 \leqslant x \leqslant \mathrm{e}$$

因此

$$E[\mathrm{e}^X] = E[Y] = \int_{-\infty}^{+\infty} xf_Y(x)\mathrm{d}x = \int_{1}^{\mathrm{e}} \mathrm{d}x = \mathrm{e} - 1$$

■

191

尽管例 2b 中求 X 的期望值的方法是适用的，但是对离散型随机变量，还有另一种方法. 以下是第 4 章命题 5.1 的直接类推.

命题 2.1 如果 X 是一个连续型随机变量，概率密度函数是 $f(x)$，则对任意实值函数 g 有

$$E[g(X)]=\int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x$$

在例 2b 中应用命题 2.1，由于 $f(x)=1$，$0<x<1$，故

$$E[\mathrm{e}^X]=\int_0^1\mathrm{e}^x\mathrm{d}x=\mathrm{e}-1$$

与前面结果一致.

命题 2.1 的证明比离散情形复杂，下面给出 $g(x)$ 非负假设下的证明.（一般情形下的证明见理论练习 2 和 3.）为此需要下述引理.

引理 2.1 对非负随机变量 Y，有

$$E[Y]=\int_0^\infty P\{Y>y\}\mathrm{d}y$$

证明 与 Y 是连续型随机变量，密度函数为 f_Y 时，给出证明. 我们有

$$\int_0^\infty P\{Y>y\}\mathrm{d}y=\int_0^\infty\int_y^\infty f_Y(x)\mathrm{d}x\mathrm{d}y$$

此处用到事实 $P\{Y>y\}=\int_y^\infty f_Y(x)\mathrm{d}x$. 在上面的等式中交换积分次序得到

192

$$\int_0^\infty P\{Y>y\}\mathrm{d}y=\int_0^\infty\left(\int_0^x\mathrm{d}y\right)f_Y(x)\mathrm{d}x=\int_0^\infty xf_Y(x)\mathrm{d}x=E[Y]$$ ■

命题 2.1 的证明 对任意的 $g(x)\geqslant 0$，由引理 2.1 可得

$$\begin{aligned}E[g(X)]&=\int_0^\infty P\{g(X)>y\}\mathrm{d}y=\int_0^\infty\int_{x:g(x)>y}f(x)\mathrm{d}x\mathrm{d}y\\&=\int_{x:g(x)>0}\int_0^{g(x)}\mathrm{d}yf(x)\mathrm{d}x=\int_{x:g(x)>0}g(x)f(x)\mathrm{d}x\end{aligned}$$

这就完成了证明. ■

例 2c 一根长度为 1 的木棍被在(0,1)上服从均匀分布的点 U 所截，求包含点 p 的那段木棍的期望长度，其中 $0\leqslant p\leqslant 1$.

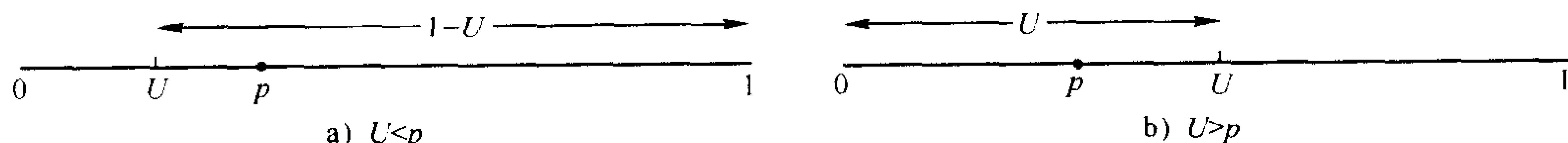
解 令 $L_p(U)$ 表示包含点 p 的那段木棍的长度，注意到（见图 5-2）

$$L_p(U)=\begin{cases}1-U & U<p\\U & U>p\end{cases}$$

因此由命题 2.1 得到

$$\begin{aligned}E[L_p(U)]&=\int_0^1L_p(u)\mathrm{d}u=\int_0^p(1-u)\mathrm{d}u+\int_p^1u\mathrm{d}u\\&=\frac{1}{2}-\frac{(1-p)^2}{2}+\frac{1}{2}-\frac{p^2}{2}=\frac{1}{2}+p(1-p)\end{aligned}$$

当 $p=\frac{1}{2}$ 时，$p(1-p)$ 最大，即当 p 是木棍的中点时，包含 p 的那段木棍的期望长度最大.



193

图 5-2 包含点 p 的那段木棍 ■

例 2d 假如你提前 s 分钟赴约，花费是 cs，如果迟到 s 分钟，花费是 ks. 假设从现在的位置到赴约的地点所用的时间是一个随机变量，概率密度函数是 f. 若想使花费最小，确定应该离开的时间.

解 令 X 表示赴约所用时间. 若在赴约前 t 分钟离开，那么花费 $C_t(X)$ 为

$$C_t(X)=\begin{cases}c(t-X) & \text{若 } X\leqslant t\\ k(X-t) & \text{若 } X\geqslant t\end{cases}$$

因此，

$$E[C_t(X)]=\int_0^{\infty}C_t(x)f(x)\mathrm{d}x=\int_0^t c(t-x)f(x)\mathrm{d}x+\int_t^{\infty}k(x-t)f(x)\mathrm{d}x$$
$$=ct\int_0^t f(x)\mathrm{d}x-c\int_0^t xf(x)\mathrm{d}x+k\int_t^{\infty}xf(x)\mathrm{d}x-kt\int_t^{\infty}f(x)\mathrm{d}x$$

现在，可以通过计算得到使 $E[C_t(X)]$达到最小的 t 的值，求导得

$$\frac{\mathrm{d}}{\mathrm{d}t}E[C_t(X)]=ctf(t)+cF(t)-ctf(t)-ktf(t)+ktf(t)-k[1-F(t)]=(k+c)F(t)-k$$

令上式等于 0，表明在你赴约前 t^* 分钟离开得到最小期望花费，而满足

$$F(t^*)=\frac{k}{k+c}$$

类似第 4 章，可由命题2.1得到如下推论. ■

推论 2.1 *若 a，b 为常数，则有*

$$E[aX+b]=aE[X]+b$$

X 是连续型随机变量时，推论2.1的证明与离散情形相同. 唯一的不同之处是将求和变为积分，将概率质量函数变为概率密度函数. [194]

连续型随机变量的方差的定义与离散情形一致. 即若 X 的期望是 μ，那么它的方差定义为

$$\mathrm{Var}(X)=E[(X-\mu)^2]$$

与离散情形类似，可得

$$\mathrm{Var}(X)=E[X^2]-(E[X])^2$$

例 2e 求例 2a 中 X 的方差.

解 首先计算 $E[X^2]$.

$$E[X^2]=\int_{-\infty}^{+\infty}x^2f(x)\mathrm{d}x=\int_0^1 2x^3\mathrm{d}x=\frac{1}{2}$$

又由 $E[X]=\frac{2}{3}$得到

$$\mathrm{Var}(X)=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{1}{18}$$ ■

模仿离散情形的证明，可以得到对常数 a 与 b 有

$$\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X)$$

在概率论的应用中，有几类重要的连续型随机变量，下面将对它们中的部分做一些介绍.

5.3 均匀随机变量

如果一个随机变量的密度函数为

$$f(x)=\begin{cases}1 & 0<x<1\\ 0 & \text{其他}\end{cases} \tag{3-1}$$

则称它服从(0,1)上的均匀分布.

注意到式(3-1)是密度函数，因为 $f(x)\geqslant 0$ 且 $\int_{-\infty}^{\infty} f(x)\mathrm{d}x=\int_0^1 \mathrm{d}x=1$. 又由于只有当$x\in(0,1)$时 $f(x)>0$，所以 x 只能取(0,1)中的值. 同时，$f(x)$在(0,1)中取常数，所以 X 在(0,1)中取任何一个值与取其他任何值的概率是非常接近的. 为验证这一点，我们指出，对任意的 $0<a<b<1$，有

195

$$P\{a\leqslant X\leqslant b\}=\int_a^b f(x)\mathrm{d}x=b-a$$

换句话说，X 在(0,1)的任意子区间中取值的概率等于该子区间的长度.

一般地，称 X 为区间(α,β)上的均匀随机变量，如果 X 的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \text{如果}\ \alpha<x<\beta\\ 0 & \text{其他}\end{cases} \tag{3-2}$$

由于 $F(a)=\int_{-\infty}^{a} f(x)\mathrm{d}x$，从式(3-2)可得，区间$(\alpha,\beta)$上的均匀随机变量的分布函数是

$$F(a)=\begin{cases}0 & a\leqslant\alpha\\ \dfrac{a-\alpha}{\beta-\alpha} & \alpha<a<\beta\\ 1 & a\geqslant\beta\end{cases}$$

图 5-3 给出了 $f(a)$与 $F(a)$的图形.

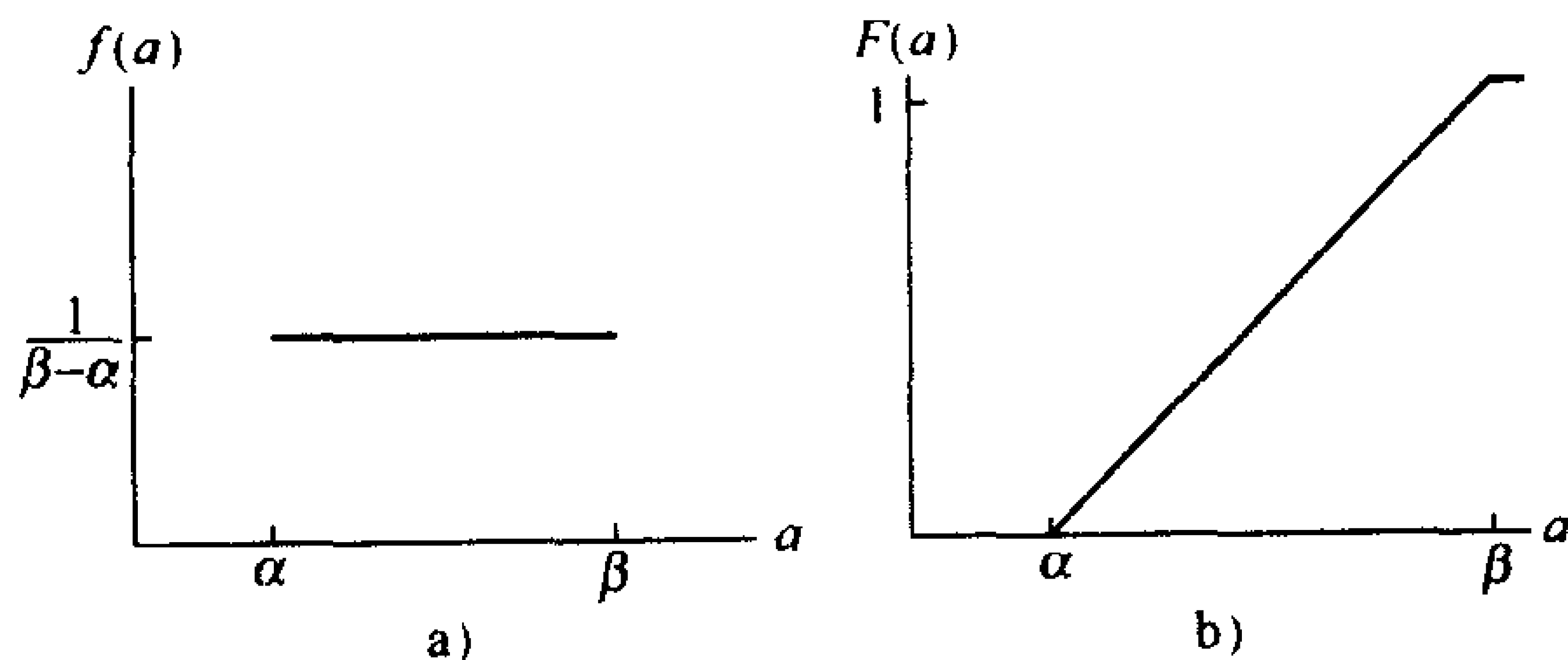


图 5-3 对于(α,β)上的均匀随机变量，a) $f(a)$和 b) $F(a)$的图形

例 3a 令 X 服从(α,β)上的均匀分布. 求(a)$E[X]$和(b)$\mathrm{Var}(X)$.

196

解 (a) $E[X]=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x=\int_\alpha^\beta \dfrac{x}{\beta-\alpha}\mathrm{d}x=\dfrac{\beta^2-\alpha^2}{2(\beta-\alpha)}=\dfrac{\beta+\alpha}{2}$

换句话说，服从某区间上均匀分布的随机变量的期望值等于该区间的中点值.

(b) 为求 $\mathrm{Var}(X)$，首先计算 $E[X^2]$.

$$E[X^2]=\int_\alpha^\beta \frac{1}{\beta-\alpha}x^2\mathrm{d}x=\frac{\beta^3-\alpha^3}{3(\beta-\alpha)}=\frac{\beta^2+\alpha\beta+\alpha^2}{3}$$

因此

$$\mathrm{Var}(X)=\frac{\beta^2+\alpha\beta+\alpha^2}{3}-\frac{(\alpha+\beta)^2}{4}=\frac{(\beta-\alpha)^2}{12}$$

因此，服从某区间上均匀分布的随机变量的方差是该区间长度平方的$\frac{1}{12}$. ■

例 3b 如果 X 服从(0,10)上的均匀分布，求下列概率：

(a) $X<3$，(b) $X>6$，(c) $3<X<8$.

解 (a) $P\{X<3\}=\int_0^3 \frac{1}{10}\mathrm{d}x=\frac{3}{10}$

(b) $P\{X>6\}=\int_6^{10} \frac{1}{10}\mathrm{d}x=\frac{4}{10}$

(c) $P\{3<X<8\}=\int_3^8 \frac{1}{10}\mathrm{d}x=\frac{1}{2}$ ■

例 3c 某公共汽车站从早上 7：00 开始每隔 15 分钟到站一辆汽车，即 7：00，7：15，7：30，7：45 等时刻有汽车到达此站. 如果一个乘客到达该站的时刻服从 7：00 到 7：30 之间的均匀分布. 求他等待下列时间的概率：

(a) 不到 5 分钟.

(b) 超过 10 分钟.

解 令 X 表示乘客 7：00 后到达该车站所等待的时间(分钟). 由于 X 是(0，30)上的均匀随机变量，所以乘客等待时间不到 5 分钟当且仅当他在 7：10 到 7：15 之间或在 7：25 到 7：30之间到达车站. 因此(a)中所求概率为 [197]

$$P\{10<X<15\}+P\{25<X<30\}=\int_{10}^{15}\frac{1}{30}\mathrm{d}x+\int_{25}^{30}\frac{1}{30}\mathrm{d}x=\frac{1}{3}$$

类似地，他等待的时间超过 10 分钟当且仅当他在 7：00 到 7：05 之间或在 7：15 到 7：20 之间到达车站，因此(b)中所求概率为

$$P\{0<X<5\}+P\{15<X<20\}=\frac{1}{3}$$ ■

下面一个例子首先由法国数学家贝特朗(L. F. Bertrand)在 1889 年提出，通常称之为贝特朗悖论. 它首先提出了一般所谓几何概率的问题.

例 3d 在某圆内作一“随机弦”，问其长大于该圆内接正三角形的边长的概率是多少？

解 上述问题的解是不确定的，这是由于“随机弦”的含义不清楚. 由于对这个词的不同理解，我们将给出此问题的两种不同解法.

第一种解法是：弦的位置可由它到圆心的距离所决定，此距离又在 0 到圆的半径 r 之间变化. 因为弦的长度大于内接正三角形的边长等价于圆心到此弦的距离小于 $r/2$，所以，若假定随机弦意味着到圆心到弦的距离 D 服从 0 到 r 之间的均匀分布，那么弦的长度大于内接正三角形的边长的概率是

$$P\left\{D<\frac{r}{2}\right\}=\frac{\frac{r}{2}}{r}=\frac{1}{2}$$

第二种解法是：考虑圆的任意一弦，过弦的一个端点作圆的切线，则弦与切线的夹角 θ 决定着弦的位置，θ 可由 0°变化到 180°(见图 5-4)；而且，弦长大于内接正三角形的边长等价于 θ 在 60°到 120°之间. 因此，若假定随机弦意味着它与切线的夹角 θ 服从 0°到 180°间的均匀分布，就得到所求的概率为

$$P\{60<\theta<120\}=\frac{120-60}{180}=\frac{1}{3}$$ ■

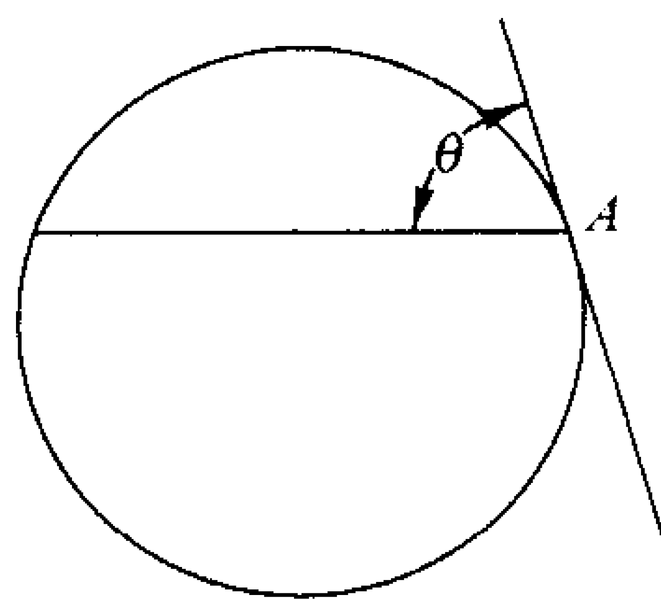


图 5-4

值得注意的是，选择不同的随机实验，可以使$\frac{1}{2}$或$\frac{1}{3}$都是此题的正确答案. 例如，如果将一半径为 r 的圆盘扔到画着一族彼此相距为 $2r$ 的平行线的桌面上，那么这族直线有且仅有一条与这个圆盘相交并形成一条弦. 圆盘的圆心到这条弦的所有距离是等可能的，从而所求弦的长度大于内接正三角形边长的概率为$\frac{1}{2}$. 另一方面，若试验是把一根针绕圆周上某固定点 A 自由旋转(见图 5-4)，则所得结果将是$\frac{1}{3}$.

5.4 正态随机变量

如果随机变量 X 的密度函数是

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}\qquad -\infty<x<+\infty$$

就称 X 是参数为 μ 与 σ^2 的正态随机变量，或简称 X 服从参数为 μ 与 σ^2 的正态分布. 这个密度函数的图形是一条关于 μ 对称的钟形曲线(见图 5-5).

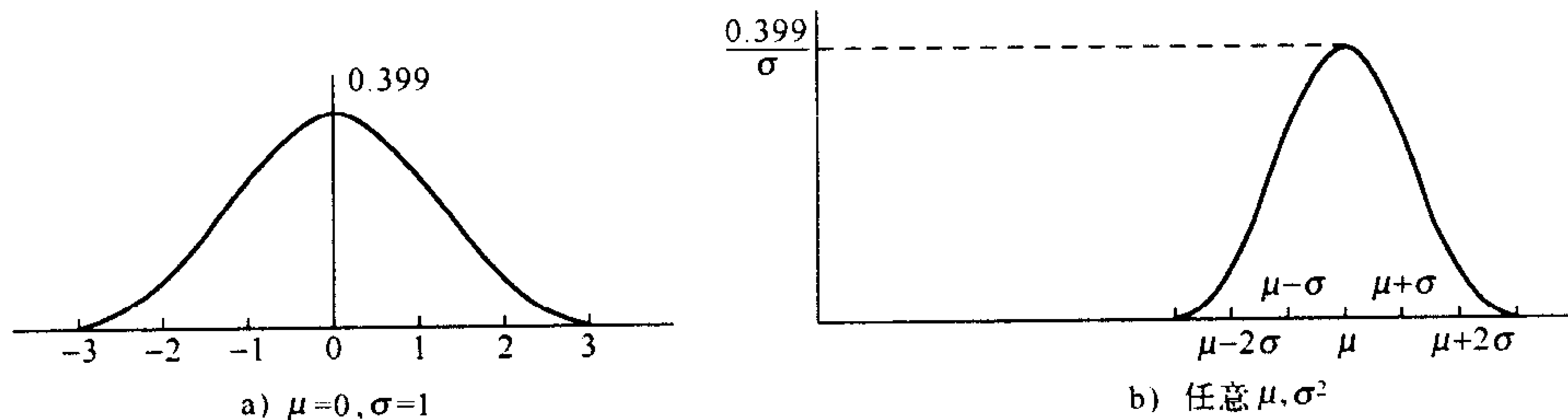


图 5-5　正态密度函数

198～199

正态分布是由法国数学家棣莫弗(Abraham DeMoivre)于 1733 年提出的，当时他用它作为二项分布随机变量当参数 n 很大时的近似概率. 后来，拉普拉斯和其他一些人推广了这一结果，现已包含在概率论著名的中心极限定理中. 极限定理将在第 8 章中讨论，它是概率论中最重要的两个结果之一[⊖]. 在实践中，诸如人的身高、气体分子向任一方向运动的速度以及测量某物理量所产生的误差等许多随机现象，都被认为是服从(至少是近似地服从)正态概率分布，而中心极限定理则给出了这一著名的经验主义见解的理论基础.

为了证明 $f(x)$的确是一个概率密度函数，需要证明

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}e^{-(x-\mu)^2/2\sigma^2}\,dx=1$$

令 $y=(x-\mu)/\sigma$，得到

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}e^{-(x-\mu)^2/2\sigma^2}\,dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-y^2/2}\,dy$$

⊖ 另一个是强大数定律.

因此必须证明

$$\int_{-\infty}^{+\infty} e^{-y^2/2} dy = \sqrt{2\pi}$$

为证此结果，令 $I = \int_{-\infty}^{+\infty} e^{-y^2/2} dy$，那么

$$I^2 = \int_{-\infty}^{+\infty} e^{-y^2/2} dy \int_{-\infty}^{+\infty} e^{-x^2/2} dx = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(y^2+x^2)/2} dy dx$$

现在通过将变量转化为极坐标计算此二重积分.（即，令 $x = r\cos\theta$，$y = r\sin\theta$，$dydx = rd\theta dr$.）于是

$$I^2 = \int_0^{\infty}\int_0^{2\pi} e^{-r^2/2} r d\theta dr = 2\pi\int_0^{\infty} re^{-r^2/2} dr = -2\pi e^{-r^2/2}\Big|_0^{\infty} = 2\pi$$

因此，$I=\sqrt{2\pi}$，结果得证.

现在证明正态随机变量的参数 μ 和 σ^2 表示它的期望值和方差. [200]

例 4a X 是参数为 μ 及 σ^2 的正态随机变量，求(a) $E[X]$及(b) $\mathrm{Var}(X)$.

解 (a) $E[X] = \dfrac{1}{\sqrt{2\pi}\sigma}\displaystyle\int_{-\infty}^{+\infty} xe^{-(x-\mu)^2/2\sigma^2} dx$

将 x 写成$(x-\mu)+\mu$ 得到

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} (x-\mu)e^{-(x-\mu)^2/2\sigma^2} dx + \mu\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} dx$$

在第一个积分中令 $y=x-\mu$ 得到

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} ye^{-y^2/2\sigma^2} dy + \mu\int_{-\infty}^{+\infty} f(x)dx$$

其中 $f(x)$是正态密度函数. 由对称性，第一个积分是 0，所以

$$E[X] = u\int_{-\infty}^{+\infty} f(x)dx = \mu$$

(b) 由于 $E[X]=\mu$，故有

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx \tag{4-1}$$

在式(4-1)中令 $y=(x-\mu)/\sigma$ 得到

$$\begin{aligned}\mathrm{Var}(X) &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} y^2 e^{-y^2/2} dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\left[-ye^{-y^2/2}\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-y^2/2} dy\right] \quad \text{由分部积分} \\ &= \sigma^2\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-y^2/2} dy = \sigma^2\end{aligned}$$

■

关于正态随机变量的一个重要事实是：若 X 服从参数为 μ 和 σ^2 的正态分布，那么 $Y=\alpha X+\beta$ 服从参数为 $\alpha\mu+\beta$ 和 $\alpha^2\sigma^2$ 的正态分布. 为了证明这个结论，假设 $\alpha>0$.（$\alpha<0$ 的证明类似.）现在，随机变量 Y 的累积分布函数 F_Y [⊖]可表示为 [201]

⊖ 当所考虑的随机变量多于一个时，用 F_Z 表示随机变量 Z 的累积分布函数. 类似地，用 f_Z 表示 Z 的密度函数.

$$F_Y(a)=P\{\alpha X+\beta\leqslant a\}=P\left\{X\leqslant\frac{a-\beta}{\alpha}\right\}=F_X\left(\frac{a-\beta}{\alpha}\right)$$

求导得到 Y 的密度函数是

$$f_Y(a)=\frac{1}{\alpha}f_X\left(\frac{a-\beta}{\alpha}\right)=\frac{1}{\sqrt{2\pi}\alpha\sigma}\exp\left\{-\left(\frac{a-\beta}{\alpha}-\mu\right)^2/2\sigma^2\right\}$$
$$=\frac{1}{\sqrt{2\pi}\alpha\sigma}\exp\{-(a-\beta-\alpha\mu)^2/2(\alpha\sigma)^2\} \qquad (4\text{-}2)$$

这表明 Y 服从均值为 $\alpha\mu+\beta$ 与方差为 $\alpha^2\sigma^2$ 的正态分布.

上述结果包含着这样一个重要的事实：若 X 服从参数为 μ 和 σ^2 的正态分布，则 $Z=\frac{X-\mu}{\sigma}$ 服从参数为 0 和 1 的正态分布. 我们称这样的随机变量 Z 有标准(或单位)正态分布.

习惯上，用 $\Phi(x)$表示一个标准正态随机变量的累积分布函数. 就是说，

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-y^2/2}\mathrm{d}y$$

对于非负的 x 值，$\Phi(x)$的值在表 5-1 中给出. 对于负的 x 值，$\Phi(x)$可由

$$\Phi(-x)=1-\Phi(x) \qquad -\infty<x<+\infty \qquad (4\text{-}3)$$

得到. 用标准正态密度的对称性可证式(4-3)，我们将它留作习题. 这个等式表明，若 Z 是一个标准正态随机变量，则

$$P\{Z\leqslant -x\}=P\{Z>x\} \qquad -\infty<x<+\infty$$

表 5-1 x 的左方标准正态曲线下的面积 $\Phi(x)$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500 0	0.504 0	0.508 0	0.512 0	0.516 0	0.519 9	0.523 9	0.527 9	0.531 9	0.535 9
0.1	0.539 8	0.543 8	0.547 8	0.551 7	0.555 7	0.559 6	0.563 6	0.567 5	0.571 4	0.575 3
0.2	0.579 3	0.583 2	0.587 1	0.591 0	0.594 8	0.598 7	0.602 6	0.606 4	0.610 3	0.614 1
0.3	0.617 9	0.621 7	0.625 5	0.629 3	0.633 1	0.636 8	0.640 6	0.644 3	0.648 0	0.651 7
0.4	0.655 4	0.659 1	0.662 8	0.666 4	0.670 0	0.673 6	0.677 2	0.680 8	0.684 4	0.687 9
0.5	0.691 5	0.695 0	0.698 5	0.701 9	0.705 4	0.708 8	0.712 3	0.715 7	0.719 0	0.722 4
0.6	0.725 7	0.729 1	0.732 4	0.735 7	0.738 9	0.742 2	0.745 4	0.748 6	0.751 7	0.754 9
0.7	0.758 0	0.761 1	0.764 2	0.767 3	0.770 4	0.773 4	0.776 4	0.779 4	0.782 3	0.785 2
0.8	0.788 1	0.791 0	0.793 9	0.796 7	0.799 5	0.802 3	0.805 1	0.807 8	0.810 6	0.813 3
0.9	0.815 9	0.818 6	0.821 2	0.823 8	0.826 4	0.828 9	0.831 5	0.834 0	0.836 5	0.838 9
1.0	0.841 3	0.843 8	0.846 1	0.848 5	0.850 8	0.853 1	0.855 4	0.857 7	0.859 9	0.862 1
1.1	0.864 3	0.866 5	0.868 6	0.870 8	0.872 9	0.874 9	0.877 0	0.879 0	0.881 0	0.883 0
1.2	0.884 9	0.886 9	0.888 8	0.890 7	0.892 5	0.894 4	0.896 2	0.898 0	0.899 7	0.901 5
1.3	0.903 2	0.904 9	0.906 6	0.908 2	0.909 9	0.911 5	0.913 1	0.914 7	0.916 2	0.917 7
1.4	0.919 2	0.920 7	0.922 2	0.923 6	0.925 1	0.926 5	0.927 9	0.929 2	0.930 6	0.931 9
1.5	0.933 2	0.934 5	0.935 7	0.937 0	0.938 2	0.939 4	0.940 6	0.941 8	0.942 9	0.944 1
1.6	0.945 2	0.946 3	0.947 4	0.948 4	0.949 5	0.950 5	0.951 5	0.952 5	0.953 5	0.954 5
1.7	0.955 4	0.956 4	0.957 3	0.958 2	0.959 1	0.959 9	0.960 8	0.961 6	0.962 5	0.963 3
1.8	0.964 1	0.964 9	0.965 6	0.966 4	0.967 1	0.967 8	0.968 6	0.969 3	0.969 9	0.970 6
1.9	0.971 3	0.971 9	0.972 6	0.973 2	0.973 8	0.974 4	0.975 0	0.975 6	0.976 1	0.976 7
2.0	0.977 2	0.977 8	0.978 3	0.978 8	0.979 3	0.979 8	0.980 3	0.980 8	0.981 2	0.981 7

（续）

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.1	0.982 1	0.982 6	0.983 0	0.983 4	0.983 8	0.984 2	0.984 6	0.985 0	0.985 4	0.985 7
2.2	0.986 1	0.986 4	0.986 8	0.987 1	0.987 5	0.987 8	0.988 1	0.988 4	0.988 7	0.989 0
2.3	0.989 3	0.989 6	0.989 8	0.990 1	0.990 4	0.990 6	0.990 9	0.991 1	0.991 3	0.991 6
2.4	0.991 8	0.992 0	0.992 2	0.992 5	0.992 7	0.992 9	0.993 1	0.993 2	0.993 4	0.993 6
2.5	0.993 8	0.994 0	0.994 1	0.994 3	0.994 5	0.994 6	0.994 8	0.994 9	0.995 1	0.995 2
2.6	0.995 3	0.995 5	0.995 6	0.995 7	0.995 9	0.996 0	0.996 1	0.996 2	0.996 3	0.996 4
2.7	0.996 5	0.996 6	0.996 7	0.996 8	0.996 9	0.997 0	0.997 1	0.997 2	0.997 3	0.997 4
2.8	0.997 4	0.997 5	0.997 6	0.997 7	0.997 7	0.997 8	0.997 9	0.997 9	0.998 0	0.998 1
2.9	0.998 1	0.998 2	0.998 2	0.998 3	0.998 4	0.998 4	0.998 5	0.998 5	0.998 6	0.998 6
3.0	0.998 7	0.998 7	0.998 7	0.998 8	0.998 8	0.998 9	0.998 9	0.998 9	0.999 0	0.999 0
3.1	0.999 0	0.999 1	0.999 1	0.999 1	0.999 2	0.999 2	0.999 2	0.999 2	0.999 3	0.999 3
3.2	0.999 3	0.999 3	0.999 4	0.999 4	0.999 4	0.999 4	0.999 4	0.999 5	0.999 5	0.999 5
3.3	0.999 5	0.999 5	0.999 5	0.999 6	0.999 6	0.999 6	0.999 6	0.999 6	0.999 6	0.999 7
3.4	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 8

由于只要 X 有以 μ 与 σ^2 为参数的正态分布，则 $Z=(X-\mu)/\sigma$ 就是标准正态随机变量. 故 X 的分布函数可以表示为

$$F_X(a)=P\{X\leqslant a\}=P\left\{\frac{X-\mu}{\sigma}\leqslant\frac{a-\mu}{\sigma}\right\}=\Phi\left(\frac{a-\mu}{\sigma}\right)$$

例 4b 若 X 是以 $\mu=3$ 与 $\sigma^2=9$ 为参数的正态随机变量，试求(a) $P\{2<X<5\}$，(b) $P\{X>0\}$，(c) $P\{|X-3|>6\}$.

解 (a) $P\{2<X<5\}=P\left\{\frac{2-3}{3}<\frac{X-3}{3}<\frac{5-3}{3}\right\}=P\left\{-\frac{1}{3}<Z<\frac{2}{3}\right\}$

$$=\Phi\left(\frac{2}{3}\right)-\Phi\left(-\frac{1}{3}\right)=\Phi\left(\frac{2}{3}\right)-\left[1-\Phi\left(\frac{1}{3}\right)\right]\approx 0.377\,9$$

202～203

(b) $P\{X>0\}=P\left\{\frac{X-3}{3}>\frac{0-3}{3}\right\}=P\{Z>-1\}=1-\Phi(-1)=\Phi(1)\approx 0.841\,3$

(c) $P\{|X-3|>6\}=P\{X>9\}+P\{X<-3\}=P\left\{\frac{X-3}{3}>\frac{9-3}{3}\right\}+P\left\{\frac{X-3}{3}<\frac{-3-3}{3}\right\}$

$$=P\{Z>2\}+P\{Z<-2\}=1-\Phi(2)+\Phi(-2)=2[1-\Phi(2)]$$

$$\approx 0.045\,6$$ ■

例 4c 进行一次考试，如果所有考生所得的分数可近似地表示为正态密度函数(换句话说，各级别考分的频率图近似地呈现正态密度的钟形曲线)，则通常认为这次考试(就合理地划分考生成绩的等级而言)是可取的. 教师经常用考试的分数去估计正态参数 μ 与 σ^2，然后把分数超过 $\mu+\sigma$ 的评为 A 等，分数在 μ 到 $\mu+\sigma$ 之间的评为 B 等，分数在 $\mu-\sigma$ 到 μ 之间的评为 C 等，分数在 $\mu-2\sigma$ 到 $\mu-\sigma$ 之间的评为 D 等，而把分数低于 $\mu-2\sigma$ 的考生评为 E 等.（有时称这种方法为“曲线上”划分等级法.）由于

$$P\{X>\mu+\sigma\}=P\left\{\frac{X-\mu}{\sigma}>1\right\}=1-\Phi(1)\approx 0.158\,7$$

$$P\{\mu<X<\mu+\sigma\}=P\left\{0<\frac{X-\mu}{\sigma}<1\right\}=\Phi(1)-\Phi(0)\approx 0.341\,3$$

$$P\{\mu-\sigma<X<\mu\}=P\left\{-1<\frac{X-\mu}{\sigma}<0\right\}=\Phi(0)-\Phi(-1)\approx 0.341\,3$$

$$P\{\mu-2\sigma<X<\mu-\sigma\}=P\left\{-2<\frac{X-\mu}{\sigma}<-1\right\}=\Phi(2)-\Phi(1)\approx 0.135\,9$$

$$P\{X<\mu-2\sigma\}=P\left\{\frac{X-\mu}{\sigma}<-2\right\}=\Phi(-2)\approx 0.022\,8$$

所以，在这次考试中获得 A 等的约占 16%，B 等的约占 34%，C 等的约占 34%，D 等的约占 14%，成绩很差的约占 2%. ■

204

例 4d 某人被控告为一个新生儿的父亲. 此案鉴定人作证时指出：母亲的怀孕期的天数（即从受孕到婴儿出生的时间）近似地服从正态分布，其参数为 $\mu=270$ 而 $\sigma^2=100$. 被告提供的证词表明，他在孩子出生前 290 天出国，而于出生前 240 天才回来. 如果被告确实是这孩子的父亲，试问那位母亲确有与证词相符的、过长或过短的怀孕期的概率是多少？

解 设 X 表示怀孕期的天数，并假定被告是这孩子的父亲，那么孩子生于与证词相符的时间内的概率是

$$\begin{aligned}P\{X>290 \text{ 或 } X<240\}&=P\{X>290\}+P\{X<240\}\\&=P\left\{\frac{X-270}{10}>2\right\}+P\left\{\frac{X-270}{10}<-3\right\}\\&=1-\Phi(2)+1-\Phi(3)\approx 0.024\,1\end{aligned}$$

■

例 4e 假设二进制信号（0 或 1）必须通过线缆从 A 地传到 B 地. 但是，数据传输受到信道噪声干扰，因此为了减小出错的概率，当信号是 1 时线缆上传输 2，信号是 0 时传输 −2. 若 $x(x=\pm 2)$ 表示 A 地发送的值，那么 B 地接收到的值是 $R(R=x+N)$，其中 N 是信道噪声干扰. 当信号传到 B 地后，接收者根据下述规则解码：

若 $R\geqslant 0.5$，解码为 1.

若 $R<0.5$，解码为 0.

由于信道噪声通常服从正态分布，我们将确定当 N 是单位正态随机变量时出错的概率.

有两种出错类型会发生：一种是信号 1 被错误地解码为 0，另一种是 0 被错误地解码为 1. 当信号是 1 并且 $2+N<0.5$ 时，会出现第一类错误；当信号是 0 并且 $-2+N\geqslant 0.5$ 时，会出现第二类错误. 因此

$$P\{\text{错误}\mid\text{信号是 }1\}=P\{N<-1.5\}=1-\Phi(1.5)\approx 0.066\,8$$

$$P\{\text{错误}\mid\text{信号是 }0\}=P\{N\geqslant 2.5\}=1-\Phi(2.5)\approx 0.006\,2$$

■

205

二项分布的正态近似

棣莫弗-拉普拉斯极限定理是概率论中的一个重要结果，它是说当 n 很大时，参数为 n 和 p 的二项随机变量近似地与均值和方差同二项分布相同的正态随机变量有相同的分布. 这个结果最初由棣莫弗于 1733 年给出 $p=\frac{1}{2}$ 时的证明，后来由拉普拉斯于 1812 年扩展到一般的 p. 这个结果在形式上是指通过先减去均值 np 然后除以标准差 $\sqrt{np(1-p)}$ 的方式“标准化”二项分布，那么标准化后的随机变量的分布函数（均值为 0、方差为 1）当 $n\to\infty$ 时收敛到标准正态分布函数.

棣莫弗-拉普拉斯极限定理 *进行 n 次独立重复试验，设每次试验成功的概率为 p，记成功的次数为 S_n，那么对任意 $a<b$，*

$$P\left\{a\leqslant\frac{S_n-np}{\sqrt{np(1-p)}}\leqslant b\right\}\to\Phi(b)-\Phi(a)\qquad n\to\infty$$

由于上述定理只是第 8 章中心极限定理的特殊情形，这里将不给出证明.

值得注意的是，对于二项分布我们已经有了两个可能的近似：当 n 充分大、p 足够小而 np 保持适当的大小时，泊松近似是一个较好的近似；另外，可以证明，当 $np(1-p)$ 较大时(见图 5-6)，正态近似是相当好的近似. [一般来说，当 $np(1-p)\geqslant10$ 时，正态近似就相当好了.]

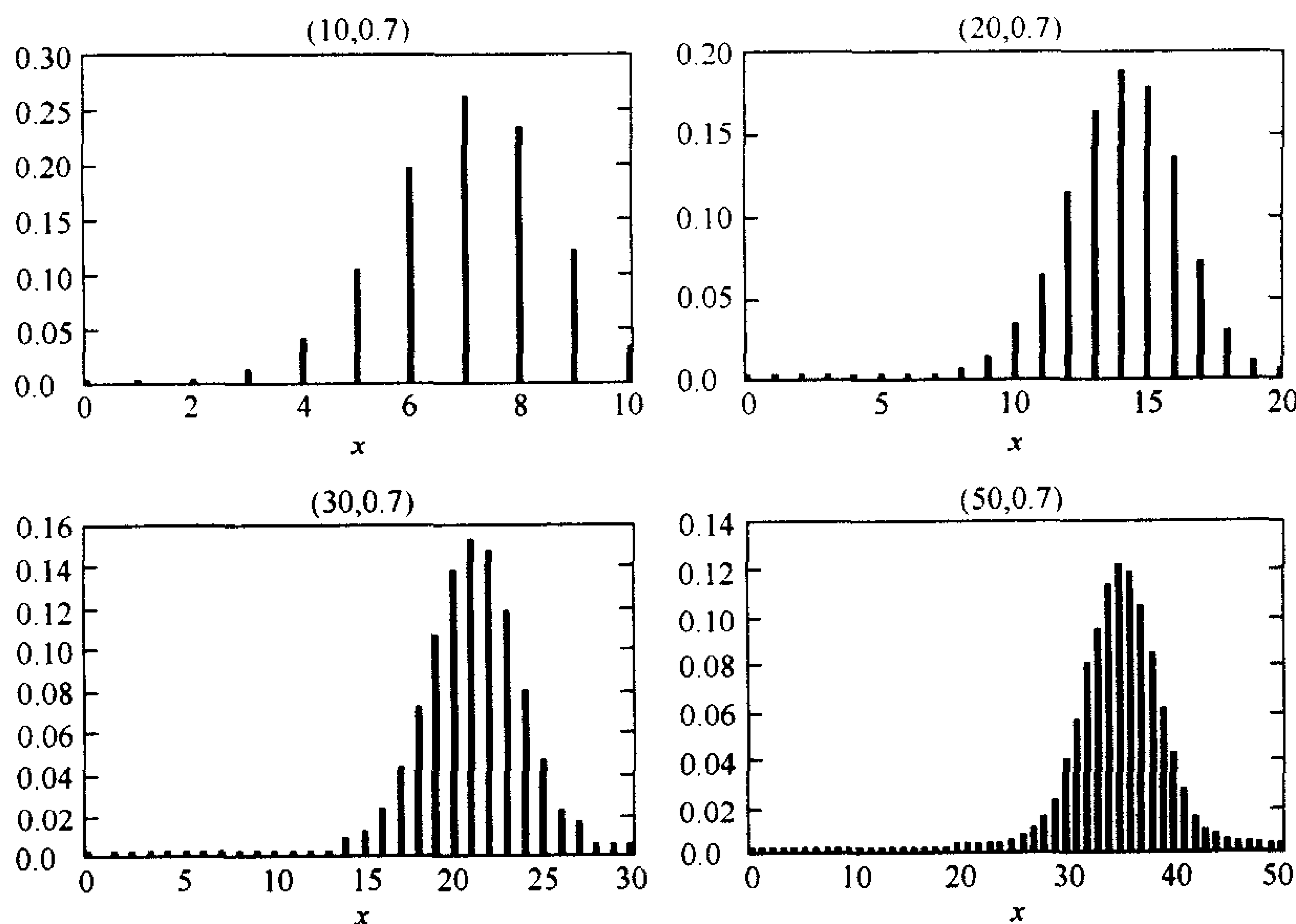


图 5-6　二项(n, p)随机变量的概率质量函数当 n 越来越大时变得越来越像正态分布

例 4f 以 X 表示抛 40 次均匀硬币出现正面的次数. 试求 $X=20$ 的概率. 运用正态近似法，再与精确解比较.

解 由于二项随机变量是离散整型的，而正态随机变量是连续型的，故要想得到较好的近似值，在应用正态近似前最好把 $P\{X=i\}$ 写成 $P\{i-1/2<X<i+1/2\}$(称之为连续性修正)，从而得到

$$P\{X=20\}=P\{19.5\leqslant X<20.5\}=P\left\{\frac{19.5-20}{\sqrt{10}}<\frac{X-20}{\sqrt{10}}<\frac{20.5-20}{\sqrt{10}}\right\}$$

$$\approx P\left\{-0.16<\frac{X-20}{\sqrt{10}}<0.16\right\}\approx\Phi(0.16)-\Phi(-0.16)\approx0.1272$$

精确解是

$$P\{X=20\}=\binom{40}{20}\left(\frac{1}{2}\right)^{40}\approx0.1254$$

■

例 4g 某学院计划招收 150 名一年级新生. 从过去的经验知道，接到录取通知的人中平均只有 30%的人入学，故学院对 450 名学生发录取通知. 试求这所学院入学新生将超过 150 名的概率.

解 以 X 表示入学新生人数，则 X 是以 $n=450$，$p=0.3$ 为参数的二项随机变量. 用连续性修正和正态近似可得

206~207

$$P\{X\geqslant 150.5\}=P\left\{\frac{X-(450)\cdot(0.3)}{\sqrt{450\cdot(0.3)\cdot(0.7)}}\geqslant\frac{150.5-(450)\cdot(0.3)}{\sqrt{450\cdot(0.3)\cdot(0.7)}}\right\}$$
$$\approx 1-\Phi(1.59)\approx 0.0559$$

因此，在第一批接到录取通知的450人中，入学者超过150名的可能性小于6%. ■

例4h 为测定能降低血液中胆固醇含量的某种食品的有效性，让100个人吃这种食品. 经充分长的时间后，化验他们的胆固醇含量. 如果至少有65%的人吃了这种食物以后胆固醇含量降低，则进行这项试验的营养学家就决定承认这种食品. 如果这种食品事实上对降低胆固醇含量不起作用，试问营养学家承认它的概率是多少?

解 假定这种食品对降低胆固醇含量不起作用，而一个人在吃了这种食品后碰巧胆固醇含量降低了的概率为$\frac{1}{2}$. 这样，若以 X 表示胆固醇含量降低的人数，则当这种食品对胆固醇含量不起作用时，营养学家承认它的概率是

$$\sum_{i=65}^{100}\binom{100}{i}\left(\frac{1}{2}\right)^{100}=P\{X\geqslant 64.5\}=P\left\{\frac{X-(100)\left(\frac{1}{2}\right)}{\sqrt{100\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}}\geqslant 2.9\right\}$$
$$\approx 1-\Phi(2.9)\approx 0.0019$$
■

例4i 52%的纽约居民赞成在公共场合禁止吸烟. 当(a) $n=11$，(b) $n=101$，(c) $n=1001$ 时，试估计在纽约随机抽取的 n 个人中，超过50%的人赞成该禁令的概率. 为了使该概率超过0.95，n 应该取多大?

解 令 N 表示纽约居民数. 为回答上述问题，必须先明白大小为 n 的随机样本是指 $\binom{N}{n}$ 个 n 个人的集合被选择的机会相同. 从而，样本中赞成禁止吸烟的人数 S_n 是一个超几何随机变量. 即 S_n 与从装有 N 个球的箱中(其中有 $0.52N$ 个是白球)抽 n 个球所得的白球数目有相同的分布. 但是由于 N 和 $0.52N$ 与样本大小 n 相比都很大，因此由超几何分布的二项分布近似得到 S_n 的分布与参数为 n 和 $p=0.52$ 的二项分布非常相近. 又由二项分布的正态近似得到

208

$$P\{S_n>0.5n\}=P\left\{\frac{S_n-0.52n}{\sqrt{n(0.52)(0.48)}}>\frac{0.5n-0.52n}{\sqrt{n(0.52)(0.48)}}\right\}$$
$$=P\left\{\frac{S_n-0.52n}{\sqrt{n(0.52)(0.48)}}>-0.04\sqrt{n}\right\}\approx\Phi(0.04\sqrt{n})$$

因此

$$P\{S_n>0.5n\}\approx\begin{cases}\Phi(0.1328)=0.5528 & \text{若 } n=11\\ \Phi(0.4020)=0.6562 & \text{若 } n=101\\ \Phi(1.2665)=0.8973 & \text{若 } n=1001\end{cases}$$

为使此概率至少为0.95，需要 $\Phi(0.04\sqrt{n})>0.95$. 因为 $\Phi(x)$ 是递增函数且 $\Phi(1.645)=0.95$，这意味着

$$0.04\sqrt{n}>1.645 \quad 或 \quad n\geqslant 1\,691.266$$

即样本大小至少为1 692. ■

关于正态分布的历史注记

正态分布由法国数学家亚伯拉罕·棣莫弗(Abraham DeMoivre,1667—1754)于1733年引入. 棣莫弗用这个分布估计与抛硬币有关的概率，称之为指数钟形曲线. 但是直到1809年，当著名德国数学家高斯(K. F. Gauss)用该方法的整数部分预测天体位置时，它的用处才真正得到体现. 因此，从此之后普遍称它为*高斯分布*.

从19世纪中期到晚期，很多统计学家开始认为大量数据有符合高斯钟形图形式的直方图. 事实上，性质良好的数据集服从此曲线被认为是"正态"的. 因此，在英国统计学家卡尔·皮尔逊(Karl Pearson)的领导下，人们开始在提到高斯曲线时，简称它为*正态曲线*.(如此多的数据集之所以服从正态曲线的部分解释将由第8章的中心极限定理给出.)

亚伯拉罕·棣莫弗

如今已经不缺乏统计顾问了，他们中的许多人在极为优雅的环境中工作. 但是，在18世纪早期，他们中的第一位在伦敦Long Acres一个又黑又脏的被称为屠夫咖啡屋的赌坊里工作. 他就是亚伯拉罕·棣莫弗——一个来自法国天主教的新教徒避难者，他能够计算出所有类型的机会游戏赌注的赔率.

正态曲线的发现者棣莫弗是一个被公认为有天分的数学家，尽管他在咖啡屋里谋生，实际上他是皇家协会的成员，并且据说是艾萨克·牛顿(Isaac Neuton)的密友.

听一下卡尔·皮尔逊想象中的棣莫弗在屠夫咖啡屋工作的情况.

"我想象棣莫弗正在咖啡屋一张肮脏的桌子上工作，旁边站着一个筋疲力尽的赌徒，艾萨克·牛顿穿过人群在角落里找到他的朋友. 这给有灵感的艺术家提供了一幅很棒的画面."

高斯

高斯(Karl Friedrich Gauss,1777—1855)，正态曲线的最早使用者之一，他是最伟大的数学家之一. 下面是著名数学史专家E. T. Bell于1954年在他的书《Men of Mathematics》中对高斯的评价. 在"The Prince of Mathematicians"(数学王子)这一章中，他写道，*"阿基米德、牛顿、高斯在最伟大的数学家中是同一级别的，就所做出的贡献来说，任何普通人都无法与他们并列. 三个人在纯数学及应用数学方面都掀起了浪潮. 阿基米德在纯数学方面比在应用数学方面更受人推崇；牛顿为他的数学发明找到了证明；高斯在纯数学及应用数学方面都取得了令人瞩目的成就."*

5.5 指数随机变量

如果某连续型随机变量的概率密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & 若\ x\geqslant 0\\ 0 & 若\ x<0\end{cases}$$

其中$\lambda>0$，则称它是以λ为参数的*指数随机变量*(或更简单地称有指数分布). 指数随机变量的累积分布函数$F(a)$由

209
~
210

$$F(a)=P\{X\leqslant a\}=\int_0^a \lambda e^{-\lambda x}\mathrm{d}x=-e^{-\lambda x}\Big|_0^a=1-e^{-\lambda a}\qquad a\geqslant 0$$

给出. 注意$F(\infty)=\int_0^{\infty}\lambda e^{-\lambda x}\mathrm{d}x=1$，当然这是它应当满足的条件. 下面将证明参数$\lambda$等于此随机变量的均值的倒数.

例 5a 令 X 是参数为 λ 的指数随机变量. 计算(a)$E[X]$和(b)$\mathrm{Var}(X)$.

解 (a) 由于密度函数由

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geqslant 0\\ 0 & x<0\end{cases}$$

给出，故可得

$$E[X]=\int_0^{\infty} x\lambda e^{-\lambda x}dx$$

由分部积分($\lambda e^{-\lambda x}dx=dv$，$u=x$)得到

$$E[X]=-xe^{-\lambda x}\Big|_0^{\infty}+\int_0^{\infty}e^{-\lambda x}dx=0-\frac{e^{-\lambda x}}{\lambda}\Big|_0^{\infty}=\frac{1}{\lambda}$$

(b) 为得到 X 的方差，首先求 $E[X^2]$.

$$E[X^2]=\int_0^{\infty}x^2\lambda e^{-\lambda x}dx$$

由分部积分($\lambda e^{-\lambda x}dx=dv$，$u=x^2$)得到

211

$$E[X^2]=-x^2e^{-\lambda x}\Big|_0^{\infty}+\int_0^{\infty}2xe^{-\lambda x}dx=0+\frac{2}{\lambda}E[X]=\frac{2}{\lambda^2}$$

因此

$$\mathrm{Var}(X)=\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2=\frac{1}{\lambda^2}$$

于是指数分布的均值是参数 λ 的倒数，方差是均值的平方. ■

实践中，到某个特定事件发生所需等待的时间往往服从指数分布. 例如，从现在开始到一次地震发生、到爆发一场新的战争、到你接到一次拨错号码的电话等所需的等待时间，都是实践中被认为有指数分布的随机变量.（关于这一点的理论解释，读者可参看 4.8 节，特别是例 8d.）

例 5b 假设打一次电话所用的时间(单位：分)是以 $\lambda=\frac{1}{10}$为参数的指数随机变量. 如果某人刚好在你前面走进公用电话亭，试求你将等待(1) 超过 10 分钟，(2) 10 分钟到 20 分钟之间的概率.

解 令 X 表示电话亭中的那个人打电话所占用的时间，所求概率为

(a) $P\{X>10\}=1-F(10)=e^{-1}\approx 0.368$

(b) $P\{10<X<20\}=F(20)-F(10)=e^{-1}-e^{-2}\approx 0.233$ ■

我们称一个非负随机变量 X 为*无记忆的*，如果对一切 s，$t\geqslant 0$ 有

$$P\{X>s+t\mid X>t\}=P\{X>s\} \tag{5-1}$$

如果设 X 是某仪器的寿命，那么式(5-1)是说：已知此仪器已使用 t 小时，它总共能工作至少 $s+t$小时的概率等于从开始使用时算起它至少能工作 s 小时的概率. 换句话说，如果仪器在时刻 t 仍"活"着，则它的剩余寿命的分布与它原来寿命的分布相同(这就是说，仪器对它已使用过的 t 小时没有记忆).

条件(5-1)等价于

212

$$\frac{P\{X>s+t, X>t\}}{P\{X>t\}}=P\{X>s\}$$

或

$$P\{X>s+t\}=P\{X>s\}P\{X>t\} \tag{5-2}$$

因为当 X 有指数分布时式(5-2)是成立的(注意到 $e^{-\lambda(s+t)}=e^{-\lambda s}e^{-\lambda t}$),所以指数分布随机变量是无记忆的.

例 5c 某邮局有两名职员值班. 假定史密斯先生走进邮局时,他发现一名职员正在为琼斯女士服务,另一名职员在接待布朗先生. 假设史密斯先生已被告知,只要琼斯或布朗一离开就接待他. 如果一名职员接待一位顾客所需要的时间服从参数为 λ 的指数分布,试问这三位顾客中,史密斯最后一个离开邮局的概率是多少?

解 将求得答案的理由陈述如下:考虑史密斯先生发现一名职员闲着的那个时刻. 此时,琼斯或布朗中有一个刚离开而另一个人的事还没办完. 但由于指数分布无记忆,故继续接待还没走的这个人(布朗或琼斯)所需的剩余服务时间仍服从参数为 λ 的指数分布. 也就是说,情况与此人在此刻刚被接待一样. 因此,由对称性可知,他在史密斯之前离开邮局的概率是 1/2. ■

可以证明,指数分布不仅是无记忆的,而且它是具有这一性质的唯一分布. 为此,设 X 是无记忆的,并令 $\bar{F}(x)=P\{X>x\}$,那么由式(5-2)可得

$$\bar{F}(s+t)=\bar{F}(s)\bar{F}(t)$$

这就是说,$\bar{F}(\cdot)$满足函数方程

$$g(s+t)=g(s)g(t)$$

但是可以证明

$$g(x)=e^{-\lambda x} \tag{5-3}$$

213

是上述函数方程的唯一右连续解[⊖],而分布函数又必须是右连续的,因此一定有

$$\bar{F}(x)=e^{-\lambda x} \quad 或 \quad F(x)=P\{X\leqslant x\}=1-e^{-\lambda x}$$

这就证明了 X 有指数分布.

例 5d 假设一辆汽车在它的蓄电池用坏之前能行使的里程数服从均值为 10 000 英里的指数分布. 如果某人决定开这辆汽车进行一次 5 000 英里的旅行,那么他不需要更换汽车的蓄电池就能完成这次旅行的概率是多少? 当上述分布不是指数分布时情况又怎样?

解 由指数分布的无记忆性可知,蓄电池的剩余寿命(以千英里计)服从以 $\lambda=\frac{1}{10}$ 为参数的指数分布. 于是所求的概率为

$$P\{剩余寿命>5\}=1-F(5)=e^{-5\lambda}=e^{-1/2}\approx 0.604$$

但是,如果寿命分布 F 不是指数分布,那么所求的概率为

$$P\{寿命>t+5\,|\,寿命>t\}=\frac{1-F(t+5)}{1-F(t)}$$

其中 t 是这次旅行开始前蓄电池已使用过的英里数. 因此,在这种情况下,计算所求的概率之前还必须知道汽车已经行驶的路程 t. ■

⊖ 函数方程的解(5-3)可以这样求得:若 $g(s+t)=g(s)g(t)$,则

$$g\left(\frac{2}{n}\right)=g\left(\frac{1}{n}+\frac{1}{n}\right)=g^2\left(\frac{1}{n}\right)$$

反复用上式可得 $g\left(\frac{m}{n}\right)=g^m\left(\frac{1}{n}\right)$. 另外还有

$$g(1)=g\left(\frac{1}{n}+\frac{1}{n}+\cdots+\frac{1}{n}\right)=g^n\left(\frac{1}{n}\right) 或 g\left(\frac{1}{n}\right)=[g(1)]^{1/n}$$

从而 $g\left(\frac{m}{n}\right)=[g(1)]^{m/n}$. 由于 g 是右连续的,故可以得到 $g(x)=[g(1)]^x$. 再因 $g(1)=\left[g\left(\frac{1}{2}\right)\right]^2\geqslant 0$,我们就得到 $g(x)=e^{-\lambda x}$,其中 $\lambda=-\log[g(1)]$.

指数分布的变种是可能取正也可能取负并且其绝对值服从参数为 $\lambda(\lambda\geqslant 0)$ 的指数分布的随机变量的分布. 这样的随机变量称为拉普拉斯分布[⊖]，密度函数为

$$f(x)=\frac{1}{2}\lambda e^{-\lambda|x|}\qquad -\infty<x<+\infty$$

分布函数为

$$F(x)=\begin{cases}\dfrac{1}{2}\displaystyle\int_{-\infty}^{x}\lambda e^{\lambda x}\,dx & x<0\\ \dfrac{1}{2}\displaystyle\int_{-\infty}^{0}\lambda e^{\lambda x}\,dx+\dfrac{1}{2}\int_{0}^{x}\lambda e^{-\lambda x}\,dx & x>0\end{cases}$$

$$=\begin{cases}\dfrac{1}{2}e^{\lambda x} & x<0\\ 1-\dfrac{1}{2}e^{-\lambda x} & x>0\end{cases}$$

例 5c 考虑例 4e，其中假设二进制信号从 A 传到 B，信号是 1 时传送 2，信号是 0 时传送 -2. 但是，假设现在信道噪声 N 不是标准正态随机变量，而是参数为 $\lambda=1$ 的拉普拉斯随机变量. 再假设 R 是在 B 地收到的值，则信号解码规则如下：

若 $R\geqslant 0.5$，解码为 1.

若 $R<0.5$，解码为 0.

在此情况下，噪声是参数为 $\lambda=1$ 的拉普拉斯随机变量，出现两种类型的错误的概率为

$$P\{\text{错误}\mid\text{信号是 }1\}=P\{N<-1.5\}=\frac{1}{2}e^{-1.5}\approx 0.1116$$

$$P\{\text{错误}\mid\text{信号是 }0\}=P\{N\geqslant 2.5\}=\frac{1}{2}e^{-2.5}\approx 0.041$$

与例 4e 的结果相比，可以看到当噪声是参数为 $\lambda=1$ 的拉普拉斯随机变量时，误差概率比噪声是标准正态随机变量时高. ■

危险率函数

考虑一个正的连续型随机变量 X，认为它是某系统的寿命，分布函数是 F，密度函数是 f. F 的危险率(hazard rate，有时称为失效率，failure rate)函数 $\lambda(t)$ 定义为

$$\lambda(t)=\frac{f(t)}{\bar{F}(t)}\qquad \bar{F}=1-F$$

为了解释 $\lambda(t)$，假设系统已存活了时间 t，求它在 dt 内出故障的概率. 即，考虑

$$P\{X\in(t,\ t+dt)\mid X>t\}=\frac{P\{X\in(t,\ t+dt),\ X>t\}}{P\{X>t\}}$$

$$=\frac{P\{X\in(t,\ t+dt)\}}{P\{X>t\}}\approx\frac{f(t)}{\bar{F}(t)}dt$$

因此，$\lambda(t)$ 表示已存活 t 单位时间的系统出故障的条件概率.

现在，假设寿命分布是指数型的，那么由无记忆性得到已存活 t 年的系统的剩余寿命的分布与一个新系统相同. 因此 $\lambda(t)$ 应该是一个常数，这已得到证实，由于

$$\lambda(t)=\frac{f(t)}{\bar{F}(t)}=\frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}=\lambda$$

⊖ 有时称为双指数随机变量.

214 215

因此，指数分布的失效率函数是常数. 参数 λ 通常称为是分布的速率.

已经证实失效率函数 $\lambda(t)$ 唯一确定了分布函数 F. 为此，注意到由定义

$$\lambda(t)=\frac{\frac{\mathrm{d}}{\mathrm{d}t}F(t)}{1-F(t)}$$

两边同时积分得到

$$\log(1-F(t))=-\int_0^t\lambda(t)\mathrm{d}t+k$$

或

$$1-F(t)=\mathrm{e}^k\exp\left\{-\int_0^t\lambda(t)\mathrm{d}t\right\}$$

令 $t=0$ 得到 $k=0$，因此

$$F(t)=1-\exp\left\{-\int_0^t\lambda(t)\mathrm{d}t\right\} \tag{5-4}$$

因此正的连续型随机变量的分布函数可由其危险率函数确定. 例如，若一个随机变量有线性危险率函数，即，若

$$\lambda(t)=a+bt$$

那么其分布函数为

$$F(t)=1-\mathrm{e}^{-at-bt^2/2}$$

求导得到密度函数为

$$f(t)=(a+bt)\mathrm{e}^{-(at+bt^2/2)}\qquad t\geqslant 0$$

当 $a=0$ 时，上式称为瑞利(Rayleigh)密度函数.

例 5f 人们经常听说同年龄的吸烟者的死亡率是不吸烟者的两倍. 这意味着什么？是不是意味着不吸烟者活过给定年龄的概率是同年龄的吸烟者的两倍？

解 若 $\lambda_s(t)$ 表示 t 年龄的吸烟者的危险率函数，$\lambda_n(t)$ 表示 t 年龄的不吸烟者的危险率函数，那么上述问题等价于

$$\lambda_s(t)=2\lambda_n(t)$$

216

A 年龄的不吸烟者活到 $B(A<B)$ 年龄的概率是

$$\begin{aligned}P\{A\text{ 年龄的不吸烟者活到 }B\text{ 年龄}\}&=P\{\text{不吸烟者的寿命}>B\,|\,\text{不吸烟者的寿命}>A\}\\&=\frac{1-F_{\text{non}}(B)}{1-F_{\text{non}}(A)}\\&=\frac{\exp\left\{-\int_0^B\lambda_n(t)\mathrm{d}t\right\}}{\exp\left\{-\int_0^A\lambda_n(t)\mathrm{d}t\right\}}\qquad\text{由式(5-4)}\\&=\exp\left\{-\int_A^B\lambda_n(t)\mathrm{d}t\right\}\end{aligned}$$

同理，吸烟者的相应概率是

$$\begin{aligned}P\{A\text{ 年龄的吸烟者活到 }B\text{ 年龄}\}&=\exp\left\{-\int_A^B\lambda_s(t)\mathrm{d}t\right\}\\&=\exp\left\{-2\int_A^B\lambda_n(t)\mathrm{d}t\right\}=\left[\exp\left\{-\int_A^B\lambda_n(t)\mathrm{d}t\right\}\right]^2\end{aligned}$$

换句话说，对两个相同年龄的人，一个吸烟另一个不吸烟，吸烟者活到给定年龄的概率是不吸

烟者的相应概率的平方(不是一半). 例如，若 $\lambda_n(t)=\frac{1}{30}$，$50\leqslant t\leqslant 60$，那么 50 岁的不吸烟者活到 60 岁的概率是 $e^{-1/3}\approx 0.716\,5$，而吸烟者的相应概率是 $e^{-2/3}\approx 0.513\,4$. ■

5.6 其他连续型随机变量

5.6.1 Γ 分布

称一个随机变量有以(α, λ)(其中 $\alpha>0$，$\lambda>0$)为参数的 Γ 分布，如果它的密度函数为

$$f(x)=\begin{cases}\dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x\geqslant 0\\ 0 & x<0\end{cases}$$

217

其中 $\Gamma(\alpha)$称为 Γ 函数，定义为

$$\Gamma(\alpha)=\int_0^\infty e^{-y}y^{\alpha-1}\,dy$$

对 $\Gamma(\alpha)$分部积分得到

$$\begin{aligned}\Gamma(\alpha)&=-e^{-y}y^{\alpha-1}\Big|_0^\infty+\int_0^\infty e^{-y}(\alpha-1)y^{\alpha-2}\,dy \qquad (6\text{-}1)\\ &=(\alpha-1)\int_0^\infty e^{-y}y^{\alpha-2}\,dy=(\alpha-1)\Gamma(\alpha-1)\end{aligned}$$

对整数值 α，比如说 $\alpha=n$，反复应用式(6-1)得到

$$\Gamma(n)=(n-1)\Gamma(n-1)=(n-1)(n-2)\Gamma(n-2)=\cdots=(n-1)(n-2)\cdots 3\cdot 2\Gamma(1)$$

由于 $\Gamma(1)=\int_0^\infty e^{-x}\,dx=1$，因此得到 n 的积分值

$$\Gamma(n)=(n-1)!$$

当 α 是正整数时，比如说 $\alpha=n$，参数为(α, λ)的 Γ 分布在实际中经常出现，它表示等待 n 个事件均发生所需的时间的分布. 更明确地说，如果一系列事件发生的时间是随机的，而且满足 4.8 节中的 3 个公理，那么等待 n 个事件发生所需的时间有参数为(n, λ)的 Γ 分布. 为证此，令 T_n 表示第 n 个事件发生的时间，注意到 T_n 小于或等于 t 当且仅当到时刻 t 为止发生的事件数至少是 n. 即，令 $N(t)$等于$[0, t]$之间发生的事件数，则

$$P\{T_n\leqslant t\}=P\{N(t)\geqslant n\}=\sum_{j=n}^\infty P\{N(t)=j\}=\sum_{j=n}^\infty\frac{e^{-\lambda t}(\lambda t)^j}{j!}$$

最后一个等号成立是由于$[0, t]$之间的事件数服从参数为 λt 的泊松分布. 对上式求导得到 T_n 的密度函数如下：

218

$$\begin{aligned}f(t)&=\sum_{j=n}^\infty\frac{e^{-\lambda t}j(\lambda t)^{j-1}\lambda}{j!}-\sum_{j=n}^\infty\frac{\lambda e^{-\lambda t}(\lambda t)^j}{j!}\\ &=\sum_{j=n}^\infty\frac{\lambda e^{-\lambda t}(\lambda t)^{j-1}}{(j-1)!}-\sum_{j=n}^\infty\frac{\lambda e^{-\lambda t}(\lambda t)^j}{j!}=\frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}\end{aligned}$$

因此，T_n 是参数为(n, λ)的 Γ 分布(此分布在文献上通常称为 n—埃尔朗分布). 注意到当 $n=1$ 时，这个分布即为指数分布.

参数为 $\lambda=\frac{1}{2}$和 $\alpha=\frac{n}{2}$(n 是正整数)的 Γ 分布称为自由度为 n 的 χ_n^2(卡方)分布. 卡方分布在实际中常作为在 n 维空间中试图击中目标时的误差，其中每个坐标的误差服从正态分布. 此分布将在第 6 章中学习，在那里还将详细讨论它与正态分布之间的关系.

例 6a 令 X 是参数为 α 和 λ 的 Γ 分布，计算(a) $E[X]$;(b) $\mathrm{Var}(X)$.

解 (a) $E[X] = \frac{1}{\Gamma(\alpha)}\int_0^\infty \lambda x \mathrm{e}^{-\lambda x}(\lambda x)^{\alpha-1}\mathrm{d}x = \frac{1}{\lambda\Gamma(\alpha)}\int_0^\infty \lambda \mathrm{e}^{-\lambda x}(\lambda x)^{\alpha}\mathrm{d}x$

$$=\frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)}$$

$$=\frac{\alpha}{\lambda} \qquad \text{由式(6-1)}$$

(b) 首先计算 $E[X^2]$，得到

$$\mathrm{Var}(X)=\frac{\alpha}{\lambda^2}$$

细节留作习题. ■ 219

5.6.2 韦布尔分布

韦布尔(Weibull)分布在工程实践中有着广泛的应用. 最初，这种分布是在解释疲劳数据时提出的，但现在它的应用已扩展到许多其他的工程问题中. 特别是，它在有关生存现象的领域中有广泛的应用. 例如，当某对象适合"最弱链"模型时，此对象的寿命服从韦布尔分布. 这就是说，考虑一个由许多部件组成的产品，并假定当它的任何一部件失效时，此产品的寿命就终止. 在这样的条件下，已经证明(理论上与实践上)韦布尔分布为产品的寿命分布提供了一个很好的近似.

韦布尔分布函数形如

$$F(x)=\begin{cases}0 & x\leqslant v\\ 1-\exp\left\{-\left(\frac{x-v}{\alpha}\right)^{\beta}\right\} & x>v\end{cases} \tag{6-2}$$

累积分布函数由式(6-2)给出的随机变量称为以 v，α，β 为参数的韦布尔随机变量. 将式(6-2)两边取导数可得其密度为

$$f(x)=\begin{cases}0 & x\leqslant v\\ \frac{\beta}{\alpha}\left(\frac{x-v}{\alpha}\right)^{\beta-1}\exp\left\{-\left(\frac{x-v}{\alpha}\right)^{\beta}\right\} & x>v\end{cases}$$

5.6.3 柯西分布

如果一个随机变量的密度函数为

$$f(x)=\frac{1}{\pi}\frac{1}{1+(x-\theta)^2} \qquad -\infty<x<+\infty$$

则称这个随机变量有以 $\theta(-\infty<\theta<+\infty)$ 为参数的柯西分布.

例 6b 假设一个手电筒在距 x 轴单位距离的上方绕其中心旋转(见图 5-7). 当手电筒停止旋转时，考虑光线与 x 轴相交的点 X.(若光线不与 x 轴相交，重复此实验.)

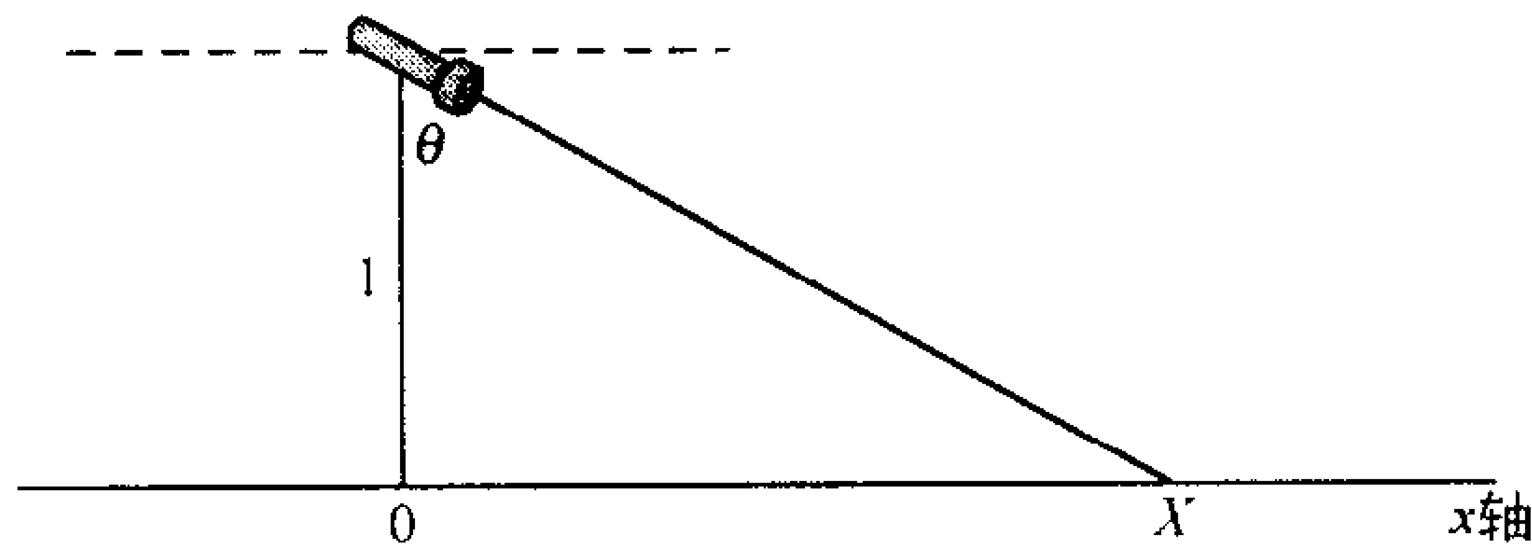


图 5-7

220

由图 5-7 可知，点 X 由手电筒与 y 轴的夹角 θ 决定，由物理知识可知 θ 服从 $-\frac{\pi}{2}$ 到 $\frac{\pi}{2}$ 之间的均匀分布. 因此 X 的分布函数由下面的式子给出：

$$F(x)=P\{X\leqslant x\}=P\{\tan\theta\leqslant x\}=P\{\theta\leqslant\tan^{-1}x\}=\frac{1}{2}+\frac{1}{\pi}\tan^{-1}x$$

最后一个等式成立是因为服从$\left(-\frac{\pi}{2},\ \frac{\pi}{2}\right)$上的均匀分布的 θ 的分布函数为

$$P\{\theta \leqslant a\}=\frac{a-(-\pi/2)}{\pi}=\frac{1}{2}+\frac{a}{\pi} \qquad -\frac{\pi}{2}<a<\frac{\pi}{2}$$

因此 X 的密度函数为

$$f(x)=\frac{\mathrm{d}}{\mathrm{d}x}F(x)=\frac{1}{\pi(1+x^2)} \qquad -\infty<x<+\infty$$

可见 X 服从柯西分布[○]. ■

5.6.4 β 分布

如果一个随机变量的密度函数是

$$f(x)=\begin{cases}\frac{1}{B(a,\ b)}x^{a-1}(1-x)^{b-1} & 0<x<1 \\ 0 & \text{其他}\end{cases}$$

其中

$$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x$$

则称这个随机变量服从 β 分布.

β 分布可用于对可能取值是有限区间$[c,\ d]$的随机现象建模，令 c 表示起点，$d-c$ 作为单位长度，则$[c,\ d]$可转化为区间$[0,1]$.

221

当 $a=b$ 时，β 密度关于$\frac{1}{2}$对称，a 增大时，密度越来越集中在$\frac{1}{2}$附近的区域(见图 5-8). 当 $b>a$ 时，密度向左偏(在此种意义下，更可能取较小的值)；当 $a>b$ 时，密度向右偏(见图 5-9).

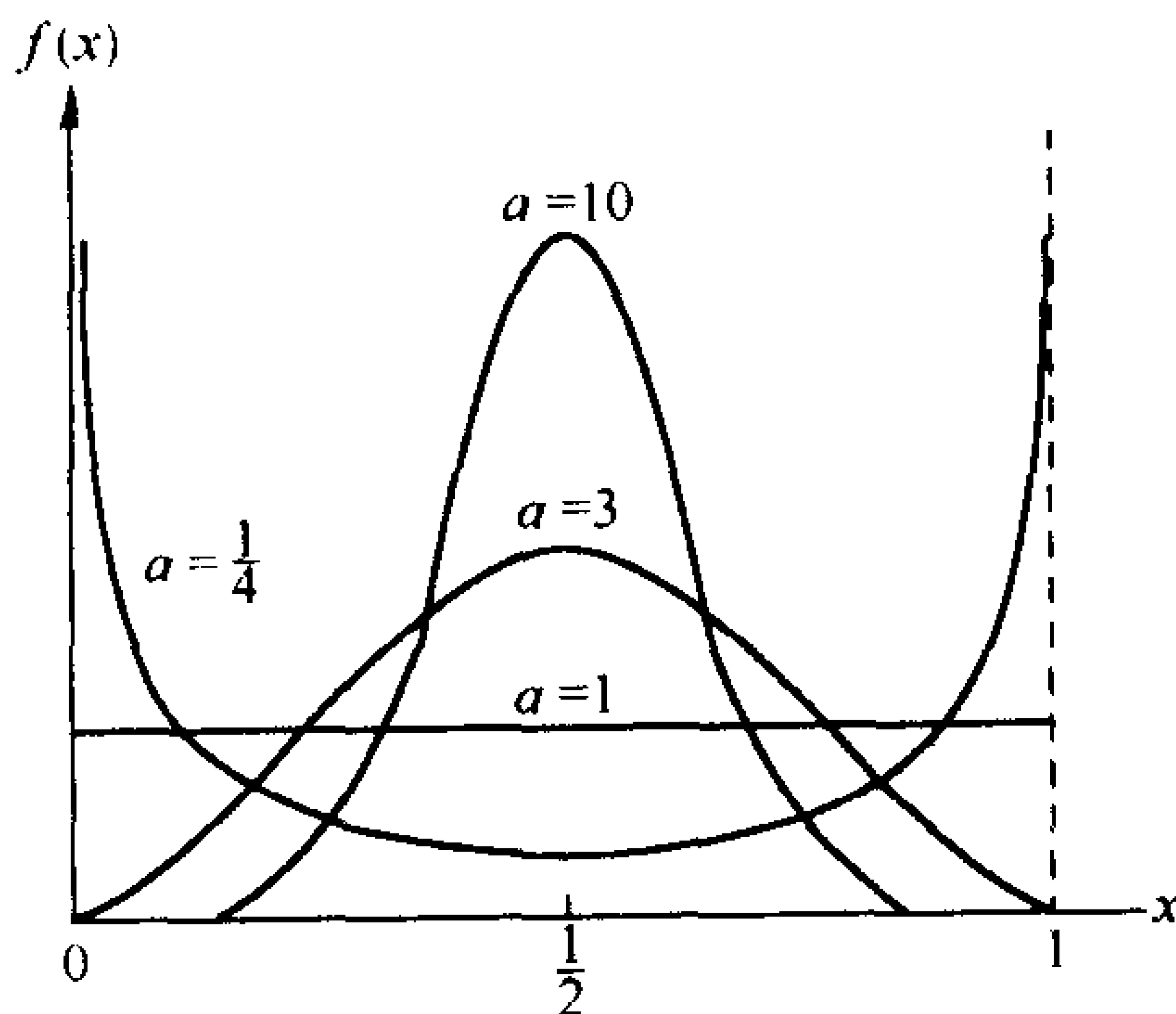


图 5-8 参数为$(a,\ b)$且 $a=b$ 时 β 分布的密度函数

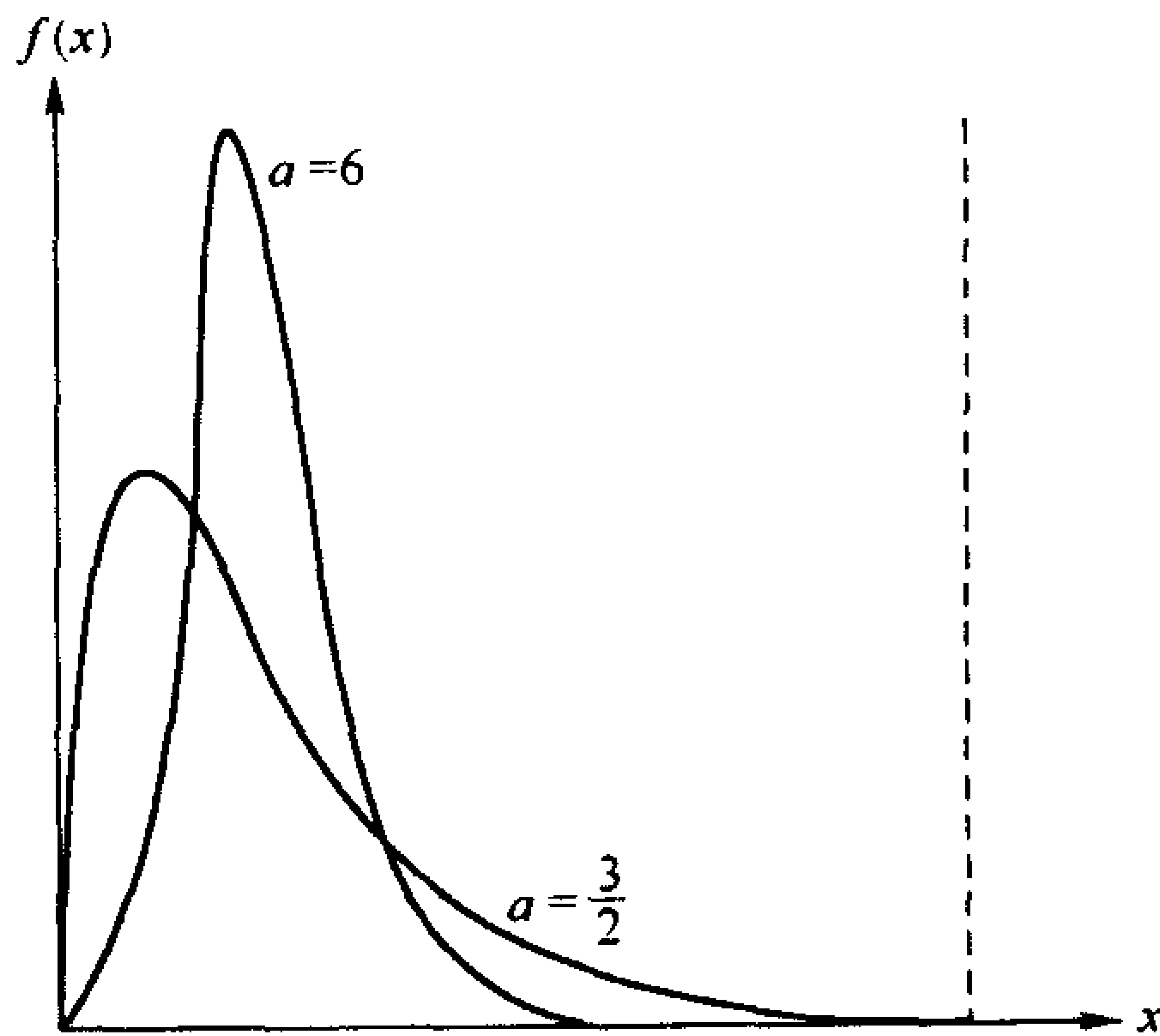


图 5-9 参数为$(a,\ b)$且 $a/(a+b)=1/20$ 时 β 分布的密度函数

○ $\frac{\mathrm{d}}{\mathrm{d}x}(\tan^{-1}x)=1/(1+x^2)$可理解如下：若 $y=\tan^{-1}x$，那么 $\tan y=x$，所以

$$1=\frac{\mathrm{d}}{\mathrm{d}x}(\tan y)=\frac{\mathrm{d}}{\mathrm{d}y}(\tan y)\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{\sin y}{\cos y}\right)\frac{\mathrm{d}y}{\mathrm{d}x}=\left(\frac{\cos^2 y+\sin^2 y}{\cos^2 y}\right)\frac{\mathrm{d}y}{\mathrm{d}x}$$

或

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\cos^2 y}{\cos^2 y+\sin^2 y}=\frac{1}{\tan^2 y+1}=\frac{1}{x^2+1}$$

可以证明在β函数与Γ函数之间存在如下关系：

$$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x$$

$$B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{6-3}$$

222

由式(6-1)及式(6-3)易见，若X是参数为a与b的β随机变量，则

$$E[X]=\frac{a}{a+b}$$

$$\mathrm{Var}(X)=\frac{ab}{(a+b)^2(a+b+1)}$$

注 式(6-3)的证明见第6章例7c.

5.7 随机变量函数的分布

经常有这种情况，一个随机变量的概率分布是已知的，要设法求出它的某个函数的分布. 例如，已知X的分布求$g(X)$的分布. 为此，必须通过有关X的集合将事件$g(X)\leqslant y$表示出来. 我们以下面的例子说明这一点.

例7a 设X服从(0，1)上的均匀分布，定义$Y=X^n$，则随机变量Y的分布用如下方法得到：对于$0\leqslant y\leqslant 1$，有

$$F_Y(y)=P\{Y\leqslant y\}=P\{X^n\leqslant y\}=P\{X\leqslant y^{1/n}\}=F_X(y^{1/n})=y^{1/n}$$

例如，Y的密度函数如下：

$$f_Y(y)=\begin{cases}\frac{1}{n}y^{1/n-1} & 0\leqslant y\leqslant 1\\ 0 & 其他\end{cases}$$ ■

例7b 若X是一个连续型随机变量，其概率密度函数为f_X，则$Y=X^2$的分布可以这样求得：对$y\geqslant 0$，有

$$F_Y(y)=P\{Y\leqslant y\}=P\{X^2\leqslant y\}=P\{-\sqrt{y}\leqslant X\leqslant\sqrt{y}\}=F_X(\sqrt{y})-F_X(-\sqrt{y})$$

223

求导得

$$f_Y(y)=\frac{1}{2\sqrt{y}}[f_X(\sqrt{y})+f_X(-\sqrt{y})]$$ ■

例7c 设X的概率密度函数为f_X，那么$Y=|X|$的密度函数可以这样求得：对$y\geqslant 0$，有

$$F_Y(y)=P\{Y\leqslant y\}=P\{|X|\leqslant y\}=P\{-y\leqslant X\leqslant y\}=F_X(y)-F_X(-y)$$

因此，求导得到

$$f_Y(y)=f_X(y)+f_X(-y)\qquad y\geqslant 0$$ ■

例7a到例7c所用的方法可用来证明定理7.1.

定理7.1 设X是一个连续型随机变量，概率密度函数是f_X. 假设$g(x)$是严格单调(递增或递减)、关于x可微(因此连续)的函数，那么随机变量$Y=g(X)$的概率密度函数为

$$f_Y(y)=\begin{cases}f_X[g^{-1}(y)]\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right| & 若存在x使y=g(x)\\ 0 & 若对任意x,y\neq g(x)\end{cases}$$

其中，$g^{-1}(y)=x$使得$g(x)=y$.

下面证明当 $g(x)$ 是递增函数时定理 7.1 成立.

证明 假设存在 x 使得 $y=g(x)$，那么，由 $Y=g(X)$ 得

$$F_Y(y)=P\{g(X)\leqslant y\}=P\{X\leqslant g^{-1}(y)\}=F_X(g^{-1}(y))$$

求导得到

224

$$f_Y(y)=f_X(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)$$

由于 $g^{-1}(y)$ 非降，从而导数非负，与定理 7.1 相符.

若对任意 x，$y\neq g(x)$，则 $F_Y(y)$ 等于 0 或 1，每一种情况下均有 $f_Y(y)=0$. ■

例 7d 令 X 是密度函数为 f 的连续型随机变量，$Y=X^n$，求 Y 的概率密度函数 f_Y.

解 令 $g(x)=x^n$，那么

$$g^{-1}(y)=y^{1/n}$$

和

$$\frac{\mathrm{d}}{\mathrm{d}y}\{g^{-1}(y)\}=\frac{1}{n}y^{1/n-1}$$

因此，由定理 7.1 可得

$$f_Y(y)=\frac{1}{n}y^{1/n-1}f(y^{1/n})$$

若 $n=2$，则有

$$f_Y(y)=\frac{1}{2\sqrt{y}}f(\sqrt{y})$$

与例 7b 的结果一致($X\geqslant 0$). ■

小结

如果存在一个非负函数 f，使得对任意的集合 B 有

$$P\{X\in B\}=\int_B f(x)\mathrm{d}x$$

则称 X 是连续型随机变量，其中 f 称为 X 的概率密度函数.

若 X 是连续型的，那么其分布函数 F 可微，并且

$$\frac{\mathrm{d}}{\mathrm{d}x}F(x)=f(x)$$

连续型随机变量 X 的均值定义为

$$E[X]=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x$$

对任意的函数 g，一个有用的恒等式为

$$E[g(X)]=\int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x$$

225

X 是离散情形时，其方差定义为

$$\mathrm{Var}(X)=E[(X-E[X])^2]$$

若随机变量 X 的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a\leqslant x\leqslant b\\ 0 & \text{其他}\end{cases}$$

则称 X 服从(a, b)上的均匀分布，它的均值和方差为

$$E[X]=\frac{a+b}{2} \qquad \mathrm{Var}(X)=\frac{(b-a)^2}{12}$$

若随机变量 X 的概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu)^2/2\sigma^2} \qquad -\infty<x<+\infty$$

则称 X 服从参数为 μ 和 σ^2 的正态分布.
可以证明

$$\mu=E[X] \qquad \sigma^2=\mathrm{Var}(X)$$

若 X 服从均值为 μ 和方差为 σ^2 的正态分布，那么

$$Z=\frac{X-\mu}{\sigma}$$

(均值为 0、方差为 1)称为标准正态随机变量. 有关 X 的随机变量可以转化为标准正态随机变量 Z 相应的概率，Z 的概率分布函数可由表 5-1 或在网站上得到.

参数为 n, p 的二项随机变量的概率分布函数，当 n 很大时，可由均值为 np、方差为 $np(1-p)$的正态分布逼近.

若一个随机变量的概率密度函数有下述形式：

$$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x} & x\geqslant 0\\ 0 & \text{其他}\end{cases}$$

则称它是参数为 λ 的指数型随机变量，其均值和方差为

$$E[X]=\frac{1}{\lambda} \qquad \mathrm{Var}(X)=\frac{1}{\lambda^2}$$

指数型随机变量的关键特性是无记忆性，即对正的 s, t 有

$$P\{X>s+t\mid X>t\}=P\{X>s\}$$

226

若 X 表示系统的寿命，则无记忆性是指已“活过”时间 t 的系统的剩余寿命与一个新系统的寿命的概率分布函数相同. 因此不必知道系统的年龄就可知道其剩余寿命的分布.

令 X 是非负的连续型随机变量，分布函数为 F，密度函数为 f. 函数

$$\lambda(t)=\frac{f(t)}{1-F(t)} \qquad t\geqslant 0$$

称为 F 的危险率或失效率函数. 如果认为 X 是系统的寿命，那么对很小的 $\mathrm{d}t$，$\lambda(t)\mathrm{d}t$ 近似地是已存活时间 t 的系统在时间 $\mathrm{d}t$ 内失效的概率. 若 F 是参数为 λ 的指数分布，那么

$$\lambda(t)=\lambda \qquad t\geqslant 0$$

另外，指数分布是唯一有常数失效率的分布.

若随机变量 X 的概率密度函数为

$$f(x)=\begin{cases}\dfrac{\lambda\mathrm{e}^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x\geqslant 0\\ 0 & \text{其他}\end{cases}$$

则称 X 服从参数为 α, λ 的 Γ 分布. $\Gamma(\alpha)$称为 Γ 函数，其定义为 $\Gamma(\alpha)=\int_0^\infty \mathrm{e}^{-x}x^{\alpha-1}\mathrm{d}x$. Γ 随机变量的均值和方差为

$$E[X]=\frac{\alpha}{\lambda} \qquad \mathrm{Var}(X)=\frac{\alpha}{\lambda^2}$$

若随机变量的概率密度函数为

$$f(x)=\begin{cases}\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1} & 0\leqslant x\leqslant 1\\ 0 & \text{其他}\end{cases}$$

则称这个随机变量服从参数为(a, b)的β分布. 常数$B(a, b)$定义为

$$B(a, b)=\int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x$$

此随机变量的均值和方差为

227

$$E[X]=\frac{a}{a+b} \qquad \mathrm{Var}(X)=\frac{ab}{(a+b)^2(a+b+1)}$$

习题

1. 设 X 是一个随机变量，其概率密度函数为

$$f(x)=\begin{cases}c(1-x^2) & -1<x<1\\ 0 & \text{其他}\end{cases}$$

(a) c 的值是多少？(b) X 的累积分布函数 $F(x)$是什么？

2. 某系统由一个原件和一个配件组成，系统的运行时间是一个随机变量 X. 若 X(单位：月)的密度函数为

$$f(x)=\begin{cases}Cxe^{-x/2} & x>0\\ 0 & x\leqslant 0\end{cases}$$

那么该系统至少运行 5 个月的概率是多少？

3. 考虑函数

$$f(x)=\begin{cases}C(2x-x^3) & 0<x<\frac{5}{2}\\ 0 & \text{其他}\end{cases}$$

f 可能是一个概率密度函数吗？若是，确定 C. 若

$$f(x)=\begin{cases}C(2x-x^2) & 0<x<\frac{5}{2}\\ 0 & \text{其他}\end{cases}$$

呢？

4. 某种电子装置的寿命 X(按小时测量)的概率密度函数为

$$f(x)=\begin{cases}\frac{10}{x^2} & x>10\\ 0 & x\leqslant 10\end{cases}$$

(a) 求 $P\{X>20\}$.

(b) X 的累积分布函数是什么？

(c) 6 种该类型的装置至少有 3 个运行至少 15 小时的概率是多少？你进行了什么样的假设？

5. 每周给加油站供一次油. 若它每周的售油量(单位：千加仑)是一个随机变量，概率密度函数为

$$f(x)=\begin{cases}5(1-x)^4 & 0<x<1\\ 0 & \text{其他}\end{cases}$$

则油箱的容积应该多大才能使得某周所供给的油被耗完的概率是 0.01？

6. 计算 $E[X]$，若 X 的密度函数给定如下：

(a) $f(x)=\begin{cases}\frac{1}{4}xe^{-x/2} & x>0\\ 0 & \text{其他}\end{cases}$；(b) $f(x)=\begin{cases}c(1-x^2) & -1<x<1\\ 0 & \text{其他}\end{cases}$；(c) $f(x)=\begin{cases}\frac{5}{x^2} & x>5\\ 0 & x\leqslant 5\end{cases}$.

228

7. 设 X 的密度函数为

$$f(x)=\begin{cases}a+bx^2 & 0\leqslant x\leqslant 1\\ 0 & \text{其他}\end{cases}$$

若 $E[X]=\frac{3}{5}$，求 a 和 b.

8. 电子管的寿命(小时)是一个随机变量，概率密度函数给定如下：

$$f(x)=xe^{-x} \qquad x\geqslant 0$$

计算该电子管的期望寿命.

9. 考虑第 4 章中的例 5b，但是现在假设季节需求量是一个连续型随机变量，概率密度函数为 f，则最优存储量 s^* 满足

$$F(s^*)=\frac{b}{b+\ell}$$

其中 b 是每单位销售量的纯利润，ℓ是单位未售出量的净损失，F 是季节需求量的累积分布函数.

10. 开往 A 地的火车从上午 7 点钟开始每隔 15 分钟到站一次，开往 B 地的火车从上午 7：05 开始每隔 15 分钟到站一次.
    (a) 若某乘客到达该站的时间服从上午 7：00～8：00 间的均匀分布，并且乘坐第一列到站的火车离开，那么他到达 A 地的时间比是多少？
    (b) 若该乘客到站的时间服从上午 7：10～8：10 间的均匀分布呢？

11. 在长度为 L 的线段上随机选定一点. 解释这句话的意思并求出短的一段与长的一段的比小于$\frac{1}{4}$的概率.

12. 公共汽车来往于相距 100 英里的 A、B 两个城市. 若汽车抛锚了，抛锚地点距 A 城的距离服从(0，100)上的均匀分布. 在 A、B 及线路中点均有一个汽车服务站. 据说三个服务站分别位于距 A 地 25、50、75 英里的地方是最有效的. 你同意吗？为什么？

13. 你 10：00 到达公共汽车站，已知公共汽车到站的时间服从 10：00 到 10：30 间的均匀分布.
    (a) 你必须等 10 分钟以上的概率是多少？
    (b) 若 10：15 时公共汽车仍未到，你至少还要再等 10 分钟的概率是多少？

14. 设 X 是(0，1)间的均匀随机变量. 用命题 2.1 计算 $E[X^n]$，然后用期望的定义核对结果.

15. 设 X 是参数为 $\mu=10$，$\sigma^2=36$ 的正态随机变量，计算
    (a) $P\{X>5\}$；(b) $P\{4<X<16\}$；(c) $P\{X<8\}$；(d) $P\{X<20\}$；(e) $P\{X>16\}$.

16. 某地的年降雨量(英寸[⊖]) 服从参数为 $\mu=40$，$\sigma=4$ 的正态分布. 从今年开始，连续 10 年的降雨量不超过 50 英寸的概率是多少？你进行了怎样的假设？

229

17. 某人进行打靶练习，若子弹离靶心的距离在 1 英寸以内，得 10 分；若在 1 至 3 英寸之间，得 5 分；在 3 至 5 英寸之间，得 3 分. 若子弹离靶心的距离服从 0 到 10 之间的均匀分布，求得分的期望值.

18. 假设 X 是均值为 5 的正态随机变量，若 $P\{X>9\}=0.2$，$\mathrm{Var}(X)$的近似值是多少？

19. 令 X 是均值为 12、方差为 4 的正态随机变量. 求 c 的值使得 $P\{X>c\}=0.10$.

20. 若社会上 65%的人赞成提高学费的提议，估计 100 人的随机样本在下列情况下的概率.

⊖ 英寸的单位符号为 in，1 in=0.025 4 m. ——编辑注

(a) 至少 50 人赞成提议.

(b) 60 到 70 人赞成提议.

(c) 少于 75 人赞成提议.

21. 假设 25 岁的男人的身高(单位：英寸)是以 $\mu=71$ 与 $\sigma^2=6.25$ 为参数的正态随机变量. 在 25 岁的男人中，身高超过 6 英尺[⊖]2 英寸的占多大比例？身高够 6 英尺的人中超过 6 英尺 2 英寸的又占多大比例？

22. 某硬铝锻件的槽的宽度(单位：英寸)有以 $\mu=0.9000$ 与 $\sigma=0.0030$ 为参数的正态分布. 规定的限度是 0.9000 ± 0.0050.

(a) 这种锻件中废品率是多少？

(b) 如果这个随机变量有以 $\mu=0.9000$ 与 σ 为参数的正态分布，为使 100 个产品中废品不多于 1 个，可允许的最大 σ 值是多少？

23. 独立掷均匀的骰子 1 000 次. 计算数字 6 出现 150～200 次的近似概率. 若数字 6 出现了 200 次，计算数字 5 出现次数少于 150 的概率.

24. 某半导体生产者生产的交互式计算机芯片的寿命服从参数为 $\mu=1.4\times10^6$，$\sigma=3\times10^5$(单位:小时)的正态分布. 一批 100 个芯片中至少含有 20 个寿命小于 1.8×10^6 小时的芯片的概率是多少？

25. 某生产者生产的每个系统独立地以 0.95 的概率被接受，估计将要生产的 150 个系统中最多有 10 个不被接受的概率.

26. 某工厂生产两种硬币：均匀的和偏重的，每个偏重硬币出现正面的概率为 55%. 我们现有这个厂生产的一枚硬币，但不知道它是均匀的还是偏重的. 为确定这枚硬币的类型，我们进行以下统计试验：将此硬币抛 1 000 次，如果正面出现了 525 次或更多，则认为它是偏重的. 如果正面出现的次数少于 525 次，则认为它是均匀的. 假定这枚硬币事实上是均匀的，我们将得到错误结论的概率是多少？若硬币本来是偏重的呢？

27. 将一枚硬币独立地掷 10 000 次，正面出现 5 800 次. 认为这枚硬币不均匀是否合理？试说明理由.

28. 有 12%的人是左撇子. 估计一所 200 人的学校中至少有 20%的人是左撇子的概率. 说明你的假设.

29. 在一个股票价格波动模型中，如果股票现在的价格是 s，那么一个时间段后，它会以概率 p 变成 us，或是以概率 $1-p$ 变成 ds. 假设连续的波动是独立的，估计在接下来的 1 000 个时间段中至少有 30%股票价格会上升的概率. 其中 $u=1.012$，$d=0.990$，$p=0.52$.

30. 将一幅图像分割成两部分—— 一部分是黑的，一部分是白的. 从白的那部分中随机地读取一个点，所读取的数服从参数为 $\mu=4$，$\sigma^2=4$ 的正态分布. 从黑的那部分中随机地读取一个点，所读取的数服从参数为(6, 9)的正态分布. 从图像中随机地读取一个点，所读取的数为 5. 图像的分割中黑色部分所占的比例为 α，当 α 取何值时，判断一个点取自于白色部分和黑色部分时犯错误的概率是相同的？

230

31. (a) 欲在长度为 $A(A<\infty)$ 的公路上建一个消防站. 若火灾发生地点服从 $(0, A)$ 上的均匀分布，为使消防站离火灾发生地点的期望距离最小，应将消防站建在哪儿？也就是说，当 X 服从 $(0, A)$ 上的均匀分布时，选择 a，使得 $E[|X-a|]$ 最小.

(b) 现在假设公路的长度是无限的——由 0 点伸向∞. 若火灾发生地点与 0 点的距离服从参数为 λ 的指数分布，消防站应建在哪儿？也就是说，当 X 服从参数为 λ 的指数分布时，使得 $E[|X-a|]$ 最小.

32. 修理一台机器所需要的时间(小时)服从参数为 $\lambda=\frac{1}{2}$ 的指数分布.

(a) 求修理时间超过 2 小时的概率.

(b) 已知修理时间已持续 9 小时的情况下，求修理时间至少是 10 小时的条件概率.

⊖ 英尺的单位符号为 ft，1 ft=0.304 8 m. ——编辑注

33. 收音机的寿命服从参数为$\lambda=\frac{1}{8}$的指数分布. 琼斯买了一台二手收音机，求它还能工作 8 年的概率.

34. 琼斯认为汽车在报废前行驶的千英里数服从参数为$\lambda=\frac{1}{20}$的指数分布. 史密斯有一辆据他说只行驶了10 000英里的二手车. 若琼斯买了这辆车，那么他至少还能行驶 20 000 英里的概率是多少？若汽车在报废前行驶的千英里数不服从指数分布，而是服从(0，40)上的均匀分布呢？

35. 年龄为t的男性吸烟者患肺癌的危险率$\lambda(t)$为

$$\lambda(t)=0.027+0.000\,25(t-40)^2 \qquad t\geqslant 40$$

若有一个 40 岁的男性吸烟者无其他任何疾病，那么他不患肺癌活到(a)50 岁和(b)60 岁的概率分别是多少？

36. 假设某系统的寿命分布的危险率函数为$\lambda(t)=t^3$，$t>0$. 求以下概率：

(a) 该系统的寿命为 2 年.

(b) 该系统的寿命在 0.4 到 1.4 之间.

(c) 系统在已使用 1 年的情况下，还能使用 2 年.

37. 若X服从$(-1,1)$上的均匀分布，求：

(a) $P\left\{|X|>\frac{1}{2}\right\}$；(b) 随机变量$|X|$的密度函数.

38. 若Y服从(0，5)上的均匀分布，那么方程$4x^2+4xY+Y+2=0$的根均为实数的概率是多少？

39. 若X服从参数为$\lambda=1$的指数分布，计算随机变量$Y=\log X$的概率密度函数.

40. 若X服从(0，1)上的均匀分布，求$Y=e^X$的密度函数.

41. 求$R=A\sin\theta$的分布，其中A是固定的常数，θ服从$(-\pi/2,\pi/2)$上的均匀分布. 这样的随机变量R出现在弹道理论中. 若在与地面成α角处将发射物点火，那么落地点R可以表示成$R=(v^2/g)\sin 2\alpha$，其中g是重力常数，等于 980 cm/s². 231

理论练习

1. 处于平衡状态的气体分子的运动速度是一个随机变量，其概率密度函数为

$$f(x)=\begin{cases}ax^2e^{-bx^2} & x\geqslant 0\\ 0 & x<0\end{cases}$$

其中$b=m/2kT$，k，T，m分别表示玻尔兹曼(Boltzmann)常数. 绝对温度与分子量. 试将a的值用b表示出来.

2. 证明

$$E[Y]=\int_0^\infty P\{Y>y\}\mathrm{d}y-\int_0^\infty P\{Y<-y\}\mathrm{d}y$$

提示：证明

$$\int_0^\infty P\{Y<-y\}\mathrm{d}y=-\int_{-\infty}^0 xf_Y(x)\mathrm{d}x$$

$$\int_0^\infty P\{Y>y\}\mathrm{d}y=\int_0^\infty xf_Y(x)\mathrm{d}x$$

3. 若X的密度函数为f，证明

$$E[g(X)]=\int_{-\infty}^\infty g(x)f(x)\mathrm{d}x$$

提示：利用理论练习 2，首先

$$E[g(X)]=\int_0^\infty P\{g(X)>y\}\mathrm{d}y-\int_0^\infty P\{g(X)<-y\}\mathrm{d}y$$

然后利用课文中给出的 $g(X)\geqslant 0$ 时的证明.

4. 证明推论 2.1.
5. 利用对非负随机变量 Y 的如下结果：

$$E[Y]=\int_0^{\infty}P\{Y>t\}\mathrm{d}t$$

证明对非负随机变量 X 有

$$E[X^n]=\int_0^{\infty}nx^{n-1}P\{X>x\}\mathrm{d}x$$

提示：首先证明

$$E[X^n]=\int_0^{\infty}P\{X^n>t\}\mathrm{d}t$$

然后作变量替换 $t=x^n$.

6. 定义事件集 $E_a(0<a<1)$ 使其具有性质：对所有 a，有

$$P(E_a)=1,\ \text{但是}\ P\left(\bigcap_a E_a\right)=0.$$

232

提示：令 X 是(0, 1)上的均匀分布，根据 X 定义每一个 E_a.

7. SD(X)表示 X 的标准差，定义如下：

$$\mathrm{SD}(X)=\sqrt{\mathrm{Var}(X)}$$

若 X 的方差是 σ^2，求 SD($aX+b$).

8. 令 X 是在 $0\sim c$ 间取值的随机变量，即 $P\{0\leqslant X\leqslant c\}=1$. 证明

$$\mathrm{Var}(X)\leqslant\frac{c^2}{4}$$

提示：首先论证

$$E[X^2]\leqslant cE[X]$$

然后利用它证明

$$\mathrm{Var}(X)\leqslant c^2[\alpha(1-\alpha)]\quad\text{其中}\quad\alpha=\frac{E[X]}{c}$$

9. 若 Z 是标准正态随机变量，证明对 $x>0$ 有
   (a) $P\{Z>x\}=P\{Z<-x\}$；(b) $P\{|Z|>x\}=2P\{Z>x\}$；(c) $P\{|Z|<x\}=2P\{Z<x\}-1$.
10. 令 $f(x)$表示均值为 μ、方差为 σ^2 的正态随机变量的概率密度函数. 证明 $\mu-\sigma$, $\mu+\sigma$ 是此函数的拐点. 即当 $x=\mu-\sigma$ 或 $x=\mu+\sigma$ 时，$f''(x)=0$.
11. 当 X 是参数为 λ 的指数随机变量时，利用理论练习 5 求 $E[X^2]$.
12. F 是连续型随机变量 X 的分布函数，称 m 是 X 的中位数，若 $F(m)=\frac{1}{2}$. 即，一个随机变量大于或小于其中位数的可能性是相同的. 求 X 的中位数，若 X 是
    (a) (a, b)上的均匀分布；(b) 参数为 μ, σ^2 的正态分布；(c) 参数为 λ 的指数分布.
13. f 是连续型随机变量 X 的密度函数，称 x 是 X 的众数，若 $f(x)$在 x 处达到最大值. 计算理论练习 12 (a)、(b)、(c)中 X 的众数.
14. 若 X 是参数为 λ 的指数随机变量，$c>0$，证明 cX 是参数为 λ/c 的指数随机变量.
15. 当 X 服从$(0, a)$上的均匀分布时，计算 X 的危险率函数.
16. 若 X 的危险率函数是 $\lambda_X(t)$，计算 aX 的危险率函数，其中 a 是正常数.
17. 证明 Γ 密度函数的积分值为 1.
18. 若 X 是均值为 $1/\lambda$ 的指数随机变量，证明

$$E[X^k]=\frac{k!}{\lambda^k}\qquad k=1, 2, \cdots$$

提示：利用Γ密度函数计算上式.

19. 证明当 X 是参数为 α 和 λ 的 Γ 随机变量时，

$$\mathrm{Var}(X)=\frac{\alpha}{\lambda^2}$$

233

20. 证明 $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$.

提示：$\Gamma\left(\frac{1}{2}\right)=\int_0^{\infty}e^{-x}x^{-1/2}dx$. 作变量替换 $y=\sqrt{2x}$，然后结合正态分布得出结果.

21. 计算参数为(α,λ)的 Γ 随机变量的失效率函数，并证明它在 $\alpha\geqslant1$ 时递增，$\alpha\leqslant1$ 时递减.
22. 计算韦布尔随机变量的失效率函数，并证明它在 $\beta\geqslant1$ 时递增，$\beta\leqslant1$ 时递减.
23. 当 $F(\cdot)$是韦布尔分布函数时，证明 $\log(\log(1-F(x))^{-1})$与 $\log(x)$的比值的图像是斜率为 β 的直线. 并证明此分布 63.2%的观察值小于 α，假设 $v=0$.
24. 令

$$Y=\left(\frac{X-v}{\alpha}\right)^{\beta}$$

证明若 X 是参数为 v，α，β 的韦布尔随机变量，那么 Y 是参数为 $\lambda=1$ 的指数随机变量. 反之亦然.

25. 若 X 是参数为 a，b 的 β 随机变量，证明

$$E[X]=\frac{a}{a+b}\qquad \mathrm{Var}(X)=\frac{ab}{(a+b)^2(a+b+1)}$$

26. 若 X 服从(a,b)上的均匀分布，则与 X 有线性关系的随机变量中，哪一个服从(0，1)上的均匀分布？
27. 考虑参数为(a,b)的 β 分布. 证明
    (a) 当 $a>1$，$b>1$ 时，密度是单峰的(即有唯一的众数)，众数等于$(a-1)/(a+b-2)$.
    (b) 当 $a\leqslant1$，$b\leqslant1$，$a+b<2$ 时，密度或者是单峰的，以 0 或 1 为众数；或者是 U-形的，以 0 和 1 为众数.
    (c) 当 $a=1$，$b=1$ 时，[0，1]中的任意一点均为众数.
28. 设 X 是累积分布函数为 F 的连续型随机变量，定义 $Y=F(X)$，证明 Y 服从(0，1)上的均匀分布.
29. X 的密度函数是 f_X. 求随机变量 Y 的概率密度函数，其中 $Y=aX+b$.
30. 当 X 是有参数为 μ，σ^2 的正态随机变量时，求 $Y=e^X$ 的概率密度函数. 称 Y 服从参数为 μ，σ^2 的对数正态分布(由于 $\log Y$ 服从正态分布).
31. 设 X，Y 是独立随机变量，并且两者取 1，2，…，$(10)^N$ 是等可能的，其中 N 很大. 令 D 表示 X，Y 的最大公因数，且 $Q_k=P\{D=k\}$.
    (a) 试给出探索性证明 $Q_k=\frac{1}{k^2}Q_1$.

    提示：注意若 D 等于 k，k 必须能够整除 X，Y，并且 X/k，Y/k 互素.(即，它们的最大公因数是 1.)

234

    (b) 利用(a)证明

$$Q_1=P\{X\text{和}Y\text{互素}\}=\frac{1}{\sum_{k=1}^{\infty}1/k^2}$$

$\sum_1^{\infty}1/k^2=\pi^2/6$ 是著名的等式，因此 $Q_1=6/\pi^2$(在数论中这称为勒让德定理).

    (c) 现在证明

$$Q_1=\prod_{i=1}^{\infty}\left(\frac{P_i^2-1}{P_i^2}\right)$$

其中 P_i 是大于 1 的第 i 个最小的素数.

提示：若 X，Y 没有公因子，则它们是互素的，因此，由(b)可得

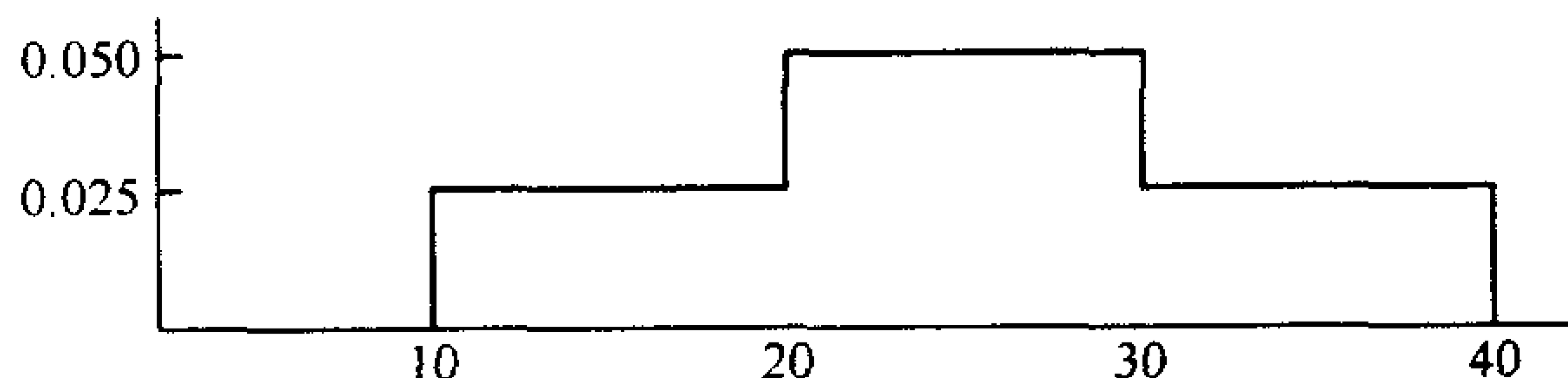
$$\prod_{i=1}^{\infty}\left(\frac{P_i^2-1}{P_i^2}\right)=\frac{6}{\pi^2}$$

此结论在第 4 章习题 11 中已不加解释地给出.（这道题与第 4 章习题 11 的关系是：若 XY 无多重素因子，那么 X，Y 互素.）

32. 当 $g(x)$是减函数时，证明定理 7.1.

自测题与练习

1. 某高中的篮球运动员在一场随机抽取的比赛中的上场时间(分钟)是一个随机变量，其概率密度函数如下图.



求运动员上场时间在下列情况下的概率.

(a) 超过 15 分钟；(b) 在 20 到 35 分钟之间；(c) 少于 30 分钟；(d) 多于 36 分钟.

2. 对某个常数 c，随机变量 X 的概率密度函数是

$$f(x)=\begin{cases}cx^n & 0<x<1\\ 0 & 其他\end{cases}$$

235

求(a) c；(b) $P\{X>x\}$，$0<x<1$.

3. 对某个常数 c，随机变量 X 的概率密度函数是

$$f(x)=\begin{cases}cx^4 & 0<x<2\\ 0 & 其他\end{cases}$$

求(a) $E[X]$；(b) $\mathrm{Var}(X)$.

4. 随机变量 X 的概率密度函数是

$$f(x)=\begin{cases}ax+bx^2 & 0<x<1\\ 0 & 其他\end{cases}$$

若 $E[X]=0.6$，求(a) $P\left\{X<\frac{1}{2}\right\}$和(b) $\mathrm{Var}(X)$.

5. 称随机变量 X 是取整数值 1，2，…，n 的均匀离散型随机变量，若

$$P\{X=i\}=\frac{1}{n}\qquad i=1,2,\cdots,n$$

对任意非负实数 x，令 $\mathrm{Int}(x)$(有时记作$[x]$)表示小于或等于 x 的最大整数值. 若 U 是(0, 1)上的均匀随机变量，证明 $X=\mathrm{Int}(nU)+1$ 是取值为 1，…，n 的均匀离散型随机变量.

6. 你的公司必须为争取某建设项目进行投标. 若你签订合同后(标最低)，打算支付另一企业 100 千美元完成该工作. 若你已确定其他参与投标的公司所投出的最大标(单位：千美元)服从(70，140)上的均匀分布，那么为了使期望利润最大，应投多大的标?

7. 为了赢得下述游戏，必须连胜三局. 游戏的胜负取决于 U，它是(0，1)上的均匀随机变量. 若 $U>0.1$，赢得第一局. 若 $U>0.2$，赢得第二局. 若 $U>0.3$，赢得第三局.

(a) 求赢得第一局的概率.

(b) 求在已经赢得第一局的条件下，赢得第二局的概率.

(c) 求在已经赢得第一局和第二局的条件下，赢得第三局的概率.

8. 随机抽取一个人，测他的IQ值，得分数近似服从均值为100、标准差为15的正态分布. 求此人的得分(a) 高于125；(b) 在90～110之间的概率.

9. 假设从你家到办公室所花费的时间服从均值为40分钟、标准差为7分钟的正态分布. 若你想以95%的可能不迟到下午1：00召开的办公室会议，你最晚应何时从家中出发?

10. 某种轮胎的寿命服从均值为34 000英里、标准差为4 000英里的正态分布.
    (a) 求此轮胎的寿命超过40 000英里的概率.
    (b) 求此轮胎的寿命在30 000英里到35 000英里之间的概率.
    (c) 在已知轮胎使用30 000英里的条件下，求它还能行驶10 000英里的条件概率.

236

11. 俄亥俄州Cleveland地区的降雨量近似地服从均值为40.2英寸、标准差为8.4英寸的正态分布.
    (a) 明年的降雨量超过44英寸的概率是多少?
    (b) 以后7年中恰有3年的降雨量超过44英寸的概率是多少?
    假定A_i是从现在算起第i年的下雨量超过44英寸的事件，证明事件$A_i(i\geqslant 1)$相互独立.

12. 下表是1992年男性和女性全职工人中年薪在不同范围的百分比的数据，

范围(美元)	女性工人的百分比	男性工人的百分比
≤9 999	8.6	4.4
10 000～19 999	38.0	21.1
20 000～24 999	19.4	15.8
25 000～49 999	29.2	41.5
≥50 000	4.8	17.2

    假设随机抽取了200个男士和200个女士，估计下述概率：
    (a) 至少70位女士收入在25 000美元或以上.
    (b) 至多60%的男士收入在25 000美元或以上.
    (c) 至少$\frac{3}{4}$的男士和至少一半的女士收入在20 000美元或以上.

13. 某银行中，顾客接受出纳员服务的时间是均值为5分钟的指数随机变量. 若你到达银行时，恰好有一位顾客在接受服务，他在接下来的4分钟仍在接受服务的概率是多少?

14. 假设随机变量X的累积分布函数给定如下：

$$F(x)=1-e^{-x^2} \qquad x>0$$

    计算(a) $P\{X>2\}$；(b) $P\{1<X<3\}$；(c) F的失效率函数；(d) $E[X]$；(e) $\mathrm{Var}(X)$.
    提示：(d)、(e)可利用理论练习5的结果.

15. 一台洗衣机工作的年限是一个随机变量，失效率函数给定如下：

$$\lambda(t)=\begin{cases}0.2 & 0<t<2\\ 0.2+0.3(t-2) & 2\leqslant t<5\\ 1.1 & t>5\end{cases}$$

    (a) 洗衣机在购买6年后仍可以工作的概率是多少?
    (b) 若洗衣机在购买6年后仍可以工作，那么它在2年内失效的概率是多少?

237

16. 一个标准的柯西随机变量的密度函数为

$$f(x)=\frac{1}{\pi(1+x^2)} \qquad -\infty<x<+\infty$$

    若X是标准的柯西随机变量，证明$1/X$也是标准的柯西随机变量.

17. 一个轮盘有 38 个槽，编号为 0,00,1～36，若你对一个特定的数字下赌注 1 元，那么，若轮盘球转到这个数字，你就得 35 元；若不转到这个数字，你就会失去 1 元. 若你连续地下赌注，估计以下概率：
(a) 下注 34 次后会赢；(b) 下注 1 000 次后会赢；(c) 下注 100 000 次后会赢；
假定每局中轮盘球都等可能地落在这 38 个数字上.

18. 盒子里有两种电池. 使用时，第 i 种电池的寿命（单位：小时）服从失效率为 $\lambda_i(i=1, 2)$ 的指数分布. 随机地抽取一种电池，抽到第 i 种的概率是 p_i，$\sum_{i=1}^{2} p_i = 1$. 若随机抽取的电池在使用 t 小时后仍可用，那么在接下来的 s 小时内它仍可用的概率是多少？

238

第6章 多个随机变量的联合分布

6.1 联合分布函数

目前为止，我们仅仅考虑单个随机变量的概率分布. 然而，我们常常需要研究涉及两个或多个随机变量的概率问题. 为了解决这类概率问题，对任意两个随机变量 X 和 Y，我们定义 X 和 Y 的**联合累积概率分布函数**(joint cumulative probability distribution function)为

$$F(a,b)=P\{X\leqslant a,\ Y\leqslant b\} \qquad -\infty<a,b<\infty$$

从 X 和 Y 的联合分布函数可以得到 X 的累积分布函数为

$$\begin{aligned}F_X(a)&=P\{X\leqslant a\}=P\{X\leqslant a,Y<\infty\}=P(\lim_{b\to\infty}\{X\leqslant a,Y\leqslant b\})\\&=\lim_{b\to\infty}P\{X\leqslant a,Y\leqslant b\}=\lim_{b\to\infty}F(a,b)\equiv F(a,\infty)\end{aligned}$$

读者应该注意，在上面的式子中，我们再次利用了概率是连续集(即事件)函数这一事实. 同理，可求得 Y 的累积分布函数为

$$F_Y(b)=P\{Y\leqslant b\}=\lim_{a\to\infty}F(a,b)\equiv F(\infty,b)$$

分布函数 F_X 和 F_Y 有时称为 X 和 Y 的**边缘分布**(marginal distribution).

理论上，X 和 Y 的联合概率问题可以用联合分布函数来解决. 例如，我们这样来求 $X>a$，$Y>b$ 的联合概率：

239

$$\begin{aligned}P\{X>a,Y>b\}&=1-P(\{X>a,Y>b\}^c)=1-P(\{X>a\}^c\cup\{Y>b\}^c)\\&=1-P(\{X\leqslant a\}\cup\{Y\leqslant b\})\\&=1-[P\{X\leqslant a\}+P\{Y\leqslant b\}-P\{X\leqslant a,Y\leqslant b\}]\\&=1-F_X(a)-F_Y(b)+F(a,b)\end{aligned} \tag{1-1}$$

式(1-1)是式(1-2)的特殊情况，证明留作习题.

$$P\{a_1<X\leqslant a_2,b_1<Y\leqslant b_2\}=F(a_2,b_2)+F(a_1,b_1)-F(a_1,b_2)-F(a_2,b_1) \tag{1-2}$$

其中 $a_1<a_2,b_1<b_2$.

当 X 和 Y 都是离散型随机变量时，X 和 Y 的**联合概率质量函数**(joint probability mass function)定义为

$$p(x,y)=P\{X=x,Y=y\}$$

从 $p(x,y)$可以求出 X 的概率质量函数为

$$p_X(x)=P\{X=x\}=\sum_{y:p(x,y)>0}p(x,y)$$

同理，

$$p_Y(y)=\sum_{x:p(x,y)>0}p(x,y)$$

例 1a 从装有 3 个红球、4 个白球、5 个蓝球的罐中随机地抽取 3 个球. 设 X 和 Y 分别代表取到的红球数和白球数，则 X 和 Y 的联合概率质量函数为 $p(i,j)=P\{X=i,Y=j\}$，分别

如下：

$$p(0,0)=\binom{5}{3}\Big/\binom{12}{3}=\frac{10}{220}$$

$$p(0,1)=\binom{4}{1}\binom{5}{2}\Big/\binom{12}{3}=\frac{40}{220}$$

$$p(0,2)=\binom{4}{2}\binom{5}{1}\Big/\binom{12}{3}=\frac{30}{220}$$

$$p(0,3)=\binom{4}{3}\Big/\binom{12}{3}=\frac{4}{220}$$

240

$$p(1,0)=\binom{3}{1}\binom{5}{2}\Big/\binom{12}{3}=\frac{30}{220}$$

$$p(1,1)=\binom{3}{1}\binom{4}{1}\binom{5}{1}\Big/\binom{12}{3}=\frac{60}{220}$$

$$p(1,2)=\binom{3}{1}\binom{4}{2}\Big/\binom{12}{3}=\frac{18}{220}$$

$$p(2,0)=\binom{3}{2}\binom{5}{1}\Big/\binom{12}{3}=\frac{15}{220}$$

$$p(2,1)=\binom{3}{2}\binom{4}{1}\Big/\binom{12}{3}=\frac{12}{220}$$

$$p(3,0)=\binom{3}{3}\Big/\binom{12}{3}=\frac{1}{220}$$

这些概率用表 6-1 表示．值得注意的是，X 的概率质量密度为行和，Y 的概率质量密度为列和．由于 X 或 Y 的概率质量函数为边缘处的值，所以通常分别称为各自的边缘概率质量函数．

表 6-1　$P\{X=i,Y=j\}$

i \ j	0	1	2	3	行和 $=P\{X=i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
列和 $=P\{Y=j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

■

例 1b 假定在一个特定的社区里 15％的家庭没有孩子，20％的家庭有 1 个孩子，35％的家庭有 2 个孩子，30％的家庭有 3 个孩子；同时进一步假设在每一个家庭中有男孩或女孩的机会均等且独立．如果从这个社区里随机地选一个家庭，B 代表男孩的数目，G 代表女孩的数目，联合分布密度函数如表 6-2 所示．

241

表 6-2 $P\{B=i,G=j\}$

i \ j	0	1	2	3	行和$=P\{B=i\}$
0	0.15	0.10	0.087 5	0.037 5	0.375 0
1	0.10	0.175	0.112 5	0	0.387 5
2	0.087 5	0.112 5	0	0	0.200 0
3	0.037 5	0	0	0	0.037 5
列和$=P\{G=j\}$	0.375 0	0.387 5	0.200 0	0.037 5	

这些概率如下所示：

$$P\{B=0,G=0\}=P\{没有孩子\}=0.15$$

$$P\{B=0,G=1\}=P\{1\text{ 个女孩且只有 }1\text{ 个孩子}\}$$

$$=P\{1\text{ 个孩子}\}P\{1\text{ 个女孩}|1\text{ 个孩子}\}=(0.20)\left(\frac{1}{2}\right)$$

$$P\{B=0,G=2\}=P\{2\text{ 个女孩且只有 }2\text{ 个孩子}\}$$

$$=P\{2\text{ 个孩子}\}P\{2\text{ 个女孩}|2\text{ 个孩子}\}=(0.35)\left(\frac{1}{2}\right)^2$$

表 6-2 中其余概率的证明留给读者. ■

称 X 和 Y 是**联合连续型的**(jointly continuous)，如果存在一个函数 $f(x,y)$，对所有的实数 x 和 y，使得对所有实数对组成的任一集合 C(即 C 是二维平面上的一个集合)具有性质：

$$P\{(X,Y)\in C\}=\iint_{(x,y)\in C} f(x,y)\mathrm{d}x\mathrm{d}y \tag{1-3}$$

函数 $f(x,y)$称为 X 与 Y 的**联合概率密度函数**(joint probability density function). 设 A,B 为任意的实数集，令 $C=\{(x,y):x\in A,y\in B\}$，则由式(1-3)可知

$$P\{X\in A,Y\in B\}=\int_B\int_A f(x,y)\mathrm{d}x\mathrm{d}y \tag{1-4}$$

因为

$$F(a,b)=P\{X\in(-\infty,a],Y\in(-\infty,b]\}=\int_{-\infty}^{b}\int_{-\infty}^{a} f(x,y)\mathrm{d}x\mathrm{d}y$$

因此，只要偏导数存在，对上式求导得

$$f(a,b)=\frac{\partial^2}{\partial a\partial b}F(a,b)$$

242

由式(1-4)可以得到联合密度函数的另外一种解释如下：当 $\mathrm{d}a$，$\mathrm{d}b$ 很小而且 $f(x,y)$在 a，b 点上连续时，

$$P\{a<X<a+\mathrm{d}a,b<Y<b+\mathrm{d}b\}=\int_b^{b+\mathrm{d}b}\int_a^{a+\mathrm{d}a} f(x,y)\mathrm{d}x\mathrm{d}y\approx f(a,b)\mathrm{d}a\mathrm{d}b$$

因此，$f(x,y)$是随机向量(X,Y)落在(a,b)附近的可能性大小的一个量度.

如果 X 和 Y 是联合连续型的，则 X 与 Y 也分别为连续型的，其概率密度函数可如下求得：

$$P\{X\in A\}=P\{X\in A,Y\in(-\infty,\infty)\}=\int_A\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x=\int_A f_X(x)\mathrm{d}x$$

其中

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y$$

是 X 的概率密度函数. 同理，Y 的概率密度函数为

$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x$$

例 1c 设 X 与 Y 的联合密度函数为

$$f(x,y)=\begin{cases}2\mathrm{e}^{-x}\mathrm{e}^{-2y} & 0<x<\infty,0<y<\infty\\ 0 & \text{其他}\end{cases}$$

计算(a)$P\{X>1,Y<1\}$，(b) $P\{X<Y\}$，(c) $P\{X<a\}$.

解

(a) $$P\{X>1,Y<1\}=\int_0^1\int_1^{\infty}2\mathrm{e}^{-x}\mathrm{e}^{-2y}\mathrm{d}x\mathrm{d}y=\int_0^1 2\mathrm{e}^{-2y}(-\mathrm{e}^{-x}\mid_1^{\infty})\mathrm{d}y$$

$$=\mathrm{e}^{-1}\int_0^1 2\mathrm{e}^{-2y}\mathrm{d}y=\mathrm{e}^{-1}(1-\mathrm{e}^{-2})$$

[243]

(b) $$P\{X<Y\}=\iint_{(x,y):x<y}2\mathrm{e}^{-x}\mathrm{e}^{-2y}\mathrm{d}x\mathrm{d}y=\int_0^{\infty}\int_0^{y}2\mathrm{e}^{-x}\mathrm{e}^{-2y}\mathrm{d}x\mathrm{d}y$$

$$=\int_0^{\infty}2\mathrm{e}^{-2y}(1-\mathrm{e}^{-y})\mathrm{d}y=\int_0^{\infty}2\mathrm{e}^{-2y}\mathrm{d}y-\int_0^{\infty}2\mathrm{e}^{-3y}\mathrm{d}y$$

$$=1-\frac{2}{3}=\frac{1}{3}$$

(c) $$P\{X<a\}=\int_0^a\int_0^{\infty}2\mathrm{e}^{-2y}\mathrm{e}^{-x}\mathrm{d}y\mathrm{d}x=\int_0^a\mathrm{e}^{-x}\mathrm{d}x=1-\mathrm{e}^{-a}$$ ■

例 1d 考虑一个半径为 R 的圆，在圆中随机地取一点满足圆中所有的区域均等可能地包含这一点(换句话说，该点在这个圆中服从均匀分布). 如果取圆的中心为原点，定义 X,Y 为所选点的坐标(见图 6-1). 由于(X,Y)等可能地落在圆中的每一点附近，所以 X 与 Y 的联合密度函数为

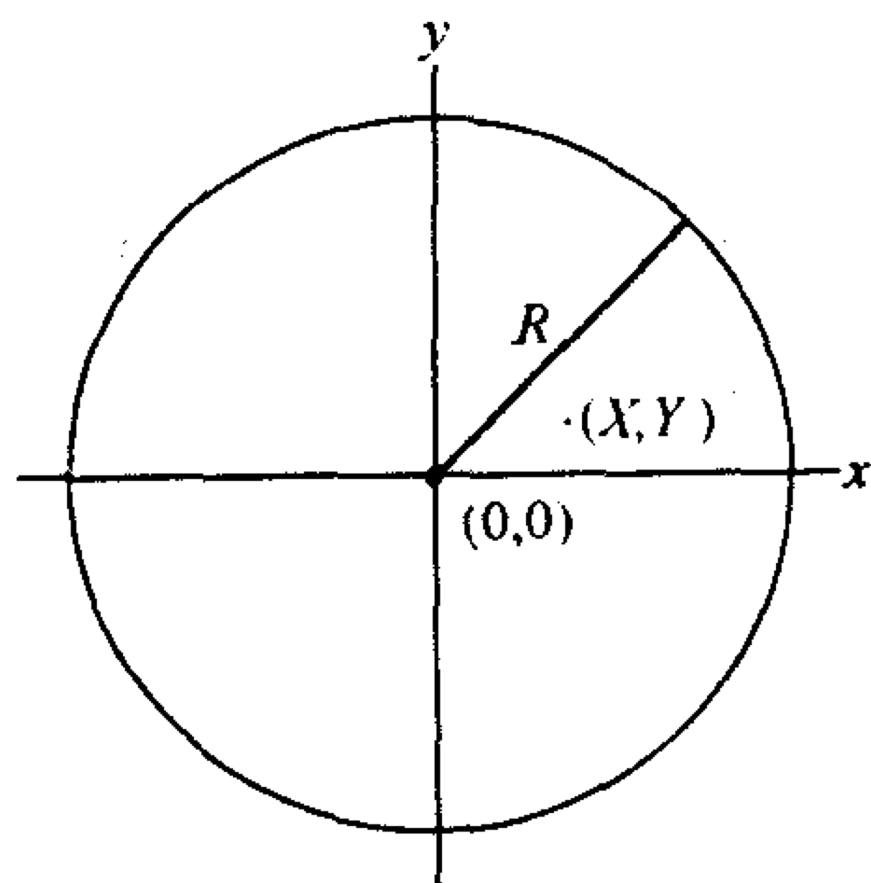


图 6-1 联合概率分布

$$f(x,y)=\begin{cases}c & \text{若 } x^2+y^2\leqslant R^2\\ 0 & \text{若 } x^2+y^2>R^2\end{cases}$$

[244]

其中 c 是某个常数.

(a) 确定 c 的值.

(b) 求 X 与 Y 的边缘密度函数.

(c) 计算所选的点到坐标原点的距离 D 不大于 a 的概率.

(d) 求 $E[D]$.

解 (a) 因为

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y\mathrm{d}x=1$$

所以

$$c\iint_{x^2+y^2\leqslant R^2}\mathrm{d}y\mathrm{d}x=1$$

使用极坐标计算$\iint_{x^2+y^2\leqslant R^2}\mathrm{d}y\mathrm{d}x$，或者更简单地，注意上述积分代表圆的面积，所以等于 πR^2. 因而，

$$c=\frac{1}{\pi R^2}$$

(b) $f_X(x)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y=\frac{1}{\pi R^2}\int_{x^2+y^2\leqslant R^2}\mathrm{d}y$

$$=\frac{1}{\pi R^2}\int_{-c}^{c}\mathrm{d}y \qquad c=\sqrt{R^2-x^2}$$

$$=\frac{2}{\pi R^2}\sqrt{R^2-x^2} \qquad x^2\leqslant R^2$$

当 $x^2>R^2$ 时，上式为 0. 根据对称性，Y 的边缘密度函数为

$$f_Y(y)=\begin{cases}\frac{2}{\pi R^2}\sqrt{R^2-y^2} & y^2\leqslant R^2\\ 0 & y^2>R^2\end{cases}$$

(c) 所选点到原点的距离 $D=\sqrt{X^2+Y^2}$ 的分布函数为：对 $0\leqslant a\leqslant R$，

$$\begin{aligned}F_D(a)&=P\{\sqrt{X^2+Y^2}\leqslant a\}=P\{X^2+Y^2\leqslant a^2\}\\&=\iint_{x^2+y^2\leqslant a^2}f(x,y)\mathrm{d}y\mathrm{d}x=\frac{1}{\pi R^2}\iint_{x^2+y^2\leqslant a^2}\mathrm{d}y\mathrm{d}x\\&=\frac{\pi a^2}{\pi R^2}=\frac{a^2}{R^2}\end{aligned}$$

这里利用 $\iint_{x^2+y^2\leqslant a^2}\mathrm{d}y\mathrm{d}x$ 是半径为 a 的圆的面积，所以等于 πa^2.

(d) 从(c)部分可以得到 D 的密度函数为

$$f_D(a)=\frac{2a}{R^2} \qquad 0\leqslant a\leqslant R$$

因此

$$E[D]=\frac{2}{R^2}\int_0^R a^2\mathrm{d}a=\frac{2R}{3}$$ ■

例 1c X 与 Y 的联合概率密度函数为

$$f(x,y)=\begin{cases}\mathrm{e}^{-(x+y)} & 0<x<\infty, 0<y<\infty\\ 0 & \text{其他}\end{cases}$$

求随机变量 X/Y 的密度函数.

解 先求随机变量 X/Y 的密度函数. 对 $a>0$,

$$\begin{aligned}F_{X/Y}(a)&=P\left\{\frac{X}{Y}\leqslant a\right\}=\iint_{x/y\leqslant a}\mathrm{e}^{-(x+y)}\mathrm{d}x\mathrm{d}y=\int_0^{\infty}\int_0^{ay}\mathrm{e}^{-(x+y)}\mathrm{d}x\mathrm{d}y\\&=\int_0^{\infty}(1-\mathrm{e}^{-ay})\mathrm{e}^{-y}\mathrm{d}y=\left\{-\mathrm{e}^{-y}+\frac{\mathrm{e}^{-(a+1)y}}{a+1}\right\}\Bigg|_0^{\infty}=1-\frac{1}{a+1}\end{aligned}$$

245 ~ 246

求微分得到 X/Y 的密度函数为 $f_{X/Y}(a)=1/(a+1)^2, 0<a<\infty$. ■

仿照二维的情形，我们用同样的方法可以定义 n 个随机变量的联合概率分布. 例如，n 个随机变量 $X_1,X_2,\cdots,X_n$ 的联合分布函数 $F(a_1,a_2,\cdots,a_n)$ 为

$$F(a_1,a_2,\cdots,a_n)=P\{X_1\leqslant a_1,X_2\leqslant a_2,\cdots,X_n\leqslant a_n\}$$

此外，如果存在一个联合概率密度函数 $f(x_1,x_2,\cdots,x_n)$，使得对 n 维空间中的任一子集 C，有

$$P\{(X_1,X_2,\cdots,X_n)\in C\}=\iint\cdots\int_{(x_1,x_2,\cdots,x_n)\in C}f(x_1,x_2,\cdots,x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$$

则 n 个随机变量是联合连续型的. 特别地，对任意 n 个实数集合 $A_1,A_2,\cdots,A_n$，有

$$P\{X_1 \in A_1, X_2 \in A_2, \cdots, X_n \in A_n\} = \int_{A_n}\int_{A_{n-1}}\cdots\int_{A_1} f(x_1, \cdots, x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$$

例 1f（多项分布） 多项分布是最重要的联合分布之一，它产生于进行 n 次独立重复实验的序列. 设每次实验的结果是 r 个可能的结果之一，且各自出现的概率分别为 $p_1, p_2, \cdots, p_r$，$\sum_{i=1}^{r} p_i = 1$. 如果令 X_i 代表 n 次实验中第 i 个结果出现的次数，则

$$P\{X_1 = n_1, X_2 = n_2, \cdots, X_r = n_r\} = \frac{n!}{n_1!\ n_2!\ \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r} \tag{1-5}$$

其中 $\sum_{i=1}^{r} n_i = n$.

这是因为，由实验的独立性假设知，对 n 次实验中第 i 个结果出现 $n_i(i=1,2,\cdots,r)$次的任一实验结果序列，其发生的概率等于 $p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$，而这样的实验结果序列共有 $n!/(n_1!\ n_2!\ \cdots n_r!)$个（$n$ 个元素，其中 n_1 个相同，另 n_2 个相同，…，另 n_r 个相同，共有 $n!/(n_1!\ n_2!\ \cdots n_r!)$ 种不同的排列），由此即得式(1-5). 以式(1-5)为联合概率质量函数的联合分布称为多项分布. 读者注意，当 $r=2$ 时，多项分布就化为二项分布.

[247]

作为多项分布的一个应用，假定将一个均匀骰子掷 9 次，则 1 出现 3 次、2 和 3 各出现 2 次、4 和 5 各出现 1 次、6 没有出现的概率等于

$$\frac{9!}{3!\ 2!\ 2!\ 1!\ 1!\ 0!}\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0 = \frac{9!}{3!\ 2!\ 2!}\left(\frac{1}{6}\right)^9$$ ■

6.2 独立随机变量

随机变量 X 和 Y 是独立的，如果对任意的实数集合 A 和 B，有

$$P\{X\in A, Y\in B\} = P\{X\in A\}P\{Y\in B\} \tag{2-1}$$

换句话说，X 和 Y 是独立的，如果对所有 A，B，事件 $E_A = \{X\in A\}$和事件 $F_B = \{Y\in B\}$独立.

利用概率论的三个公理可以证明式(2-1)成立，当且仅当对所有的 a，b，有

$$P\{X\leqslant a, Y\leqslant b\} = P\{X\leqslant a\}P\{Y\leqslant b\}$$

因此，如果 X 和 Y 的联合分布函数用 F 来表示，则 X 和 Y 独立，当且仅当对一切 a，b，有

$$F(a,b) = F_X(a)F_Y(b)$$

当 X 和 Y 是离散型随机变量时，独立性条件式(2-1)等价于对所有的 x, y，有

$$p(x,y) = p_X(x)p_Y(y) \tag{2-2}$$

实际上，如果式(2-1)满足，我们令其中的 A 和 B 分别为单点集 $A=\{x\}, B=\{y\}$，即得式(2-2). 反之，式(2-2)成立，则对任意集合 A，B，有

$$\begin{aligned} P\{X \in A, Y \in B\} &= \sum_{y\in B}\sum_{x\in A} p(x,y) = \sum_{y\in B}\sum_{x\in A} p_X(x)p_Y(y) \\ &= \sum_{y\in B} p_Y(y)\sum_{x\in A} p_X(x) = P\{Y \in B\}P\{X \in A\} \end{aligned}$$

从而式(2-1)成立.

在联合连续的情况下，条件独立性等价于对所有的 x, y，有

[248]

$$f(x,y) = f_X(x)f_Y(y)$$

因此，更一般地讲，X 和 Y 是独立的，如果已知其中的一个随机变量的值不会改变另一个随机变量的分布．随机变量不独立，则称为相关的．

例 2a 假设进行 $n+m$ 次独立实验，成功的概率均为 p．如果 X 代表前 n 次实验中成功的次数，Y 代表后 m 次实验中成功的次数，则 X 与 Y 是独立的，因为前 n 次实验中成功的次数不会影响后 m 次实验中成功次数的分布（因为假设是独立实验）．事实上，对整数 x,y，有

$$P\{X=x,Y=y\}=\binom{n}{x}p^x(1-p)^{n-x}\binom{m}{y}p^y(1-p)^{m-y} \quad 0\leqslant x\leqslant n,0\leqslant y\leqslant m$$
$$=P\{X=x\}P\{Y=y\}$$

另一方面，X 与 Z 是相关的，其中 Z 是 $n+m$ 次实验中成功的总次数．（为什么？） ■

例 2b 设在某一天里进入邮局的人数是一个参数为 λ 的泊松随机变量．试证：如果每一个进入邮局的人是男性的概率为 p，是女性的概率为 $1-p$，则进入邮局的男性人数和女性人数是分别以 λp 和 $\lambda(1-p)$ 为参数的独立泊松随机变量．

解 令 X 与 Y 分别代表进入邮局的男性人数和女性人数．利用式(2-2)来证明 X 和 Y 的独立性．为得到 $P\{X=i,Y=j\}$ 的表达式，将 $X+Y$ 作为条件如下：

$$P\{X=i,Y=j\}=P\{X=i,Y=j|X+Y=i+j\}P\{X+Y=i+j\}$$
$$+P\{X=i,Y=j|X+Y\neq i+j\}P\{X+Y\neq i+j\}$$

[读者应该注意到，上述式子是 $P(E)=P(E|F)P(F)+P(E|F^c)P(F^c)$ 的一个特例．]

由于 $P\{X=i,Y=j|X+Y\neq i+j\}$ 为 0，所以得到

$$P\{X=i,Y=j\}=P\{X=i,Y=j|X+Y=i+j\}P\{X+Y=i+j\} \tag{2-3}$$

现在，因为 $X+Y$ 代表进入邮局的总人数，故由假设可知

$$P\{X+Y=i+j\}=\mathrm{e}^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!} \tag{2-4}$$

此外，考虑 $i+j$ 个人进入邮局，因为每个进入邮局的人是男性的概率为 p，所以恰有 i 个人为男性（从而有 j 个人为女性）的概率正好是二项概率 $\binom{i+j}{i}p^i(1-p)^j$．也就是说，

249

$$P\{X=i,Y=j|X+Y=i+j\}=\binom{i+j}{i}p^i(1-p)^j \tag{2-5}$$

将式(2-4)、(2-5)代入式(2-3)，则

$$P\{X=i,Y=j\}=\binom{i+j}{i}p^i(1-p)^j\mathrm{e}^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}=\mathrm{e}^{-\lambda}\frac{(\lambda p)^i}{i!\ j!}[\lambda(1-p)]^j$$
$$=\frac{\mathrm{e}^{-\lambda p}(\lambda p)^i}{i!}\mathrm{e}^{-\lambda(1-p)}\frac{[\lambda(1-p)]^j}{j!} \tag{2-6}$$

从而

$$P\{X=i\}=\mathrm{e}^{-\lambda p}\frac{(\lambda p)^i}{i!}\sum_j\mathrm{e}^{-\lambda(1-p)}\frac{[\lambda(1-p)]^j}{j!}=\mathrm{e}^{-\lambda p}\frac{(\lambda p)^i}{i!} \tag{2-7}$$

同理可得

$$P\{Y=j\}=\mathrm{e}^{-\lambda(1-p)}\frac{[\lambda(1-p)]^j}{j!} \tag{2-8}$$

式(2-6)、(2-7)及(2-8)即是我们所要证明的结果． ■

例 2c 一个男人和一个女人约定在某地相会，假如每个人到达的时间是相互独立的，且在中午 12 时到下午 1 时之间是均匀分布，求先到者要等待 10 分钟以上的概率.

解 设男人到达的时间为 12 时 X 分，女人到达的时间是 12 时 Y 分，则 X 和 Y 为独立随机变量，服从(0,60)上的均匀分布. 要求的概率为 $P\{X+10<Y\}+P\{Y+10<X\}$，由对称性可知等于 $2P\{X+10<Y\}$，而

$$2P\{X+10<Y\}=2\iint_{x+10<y} f(x,y)\mathrm{d}x\mathrm{d}y=2\iint_{x+10<y} f_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y$$

250

$$=2\int_{10}^{60}\int_{0}^{y-10}\left(\frac{1}{60}\right)^2\mathrm{d}x\mathrm{d}y=\frac{2}{(60)^2}\int_{10}^{60}(y-10)\mathrm{d}y=\frac{25}{36}$$ ■

下一个例子是涉及几何概率的一个最古老的问题，通常称为蒲丰(Buffon)投针问题，它是 18 世纪法国博物学家蒲丰最先研究和解决的.

例 2d(蒲丰投针问题) 桌面上画有一些平行线，每两条线之间的距离都等于 D，一根长度为 $L(L\leqslant D)$的针随机地投在桌面上，试问此针与其中一条平行线相交的概率是多少(另外一种可能是这根针完全落在两条平行线之间的带中)?

解 以 X 表示这根针的中点到离它最近的一条平行线的距离，以 θ 表示这根针与 X 所在的投影线的夹角(见图 6-2)，如果图 6-2 中右方直角三角形的斜边的长度小于 $L/2$，则此针必与一条线相交. 即当

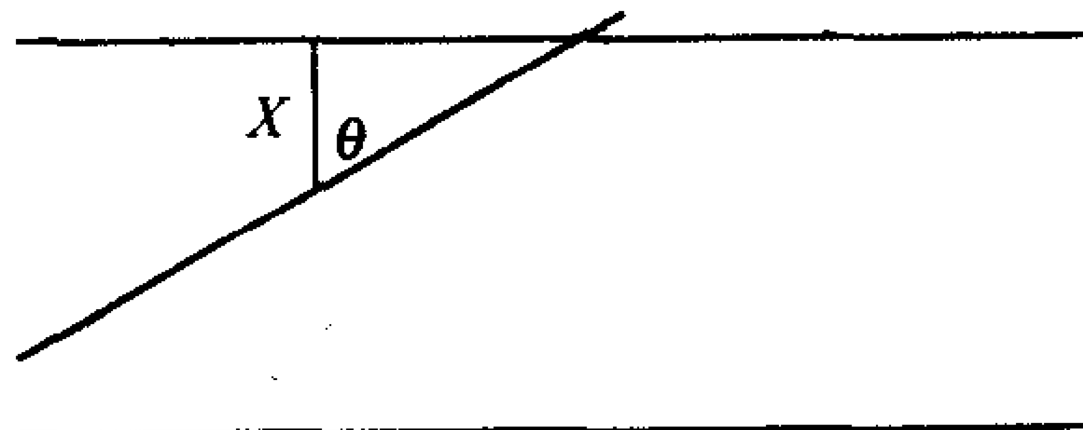


图 6-2

$$\frac{X}{\cos\theta}<\frac{L}{2} \text{ 或 } X<\frac{L}{2}\cos\theta$$

时，针与一条线相交. 因为 X 在 0 和 $D/2$ 之间变化，θ 在 0 到 $\pi/2$ 之间变化，故可设它们是独立的，并且在自己的变化范围上是均匀分布的随机变量. 因此

$$P\{X<\frac{L}{2}\cos\theta\}=\iint_{x<L/2\cos y} f_X(x)f_\theta(y)\mathrm{d}x\mathrm{d}y=\frac{4}{\pi D}\int_0^{\pi/2}\int_0^{L/2\cos y}\mathrm{d}x\mathrm{d}y$$

$$=\frac{4}{\pi D}\int_0^{\pi/2}\frac{L}{2}\cos y\mathrm{d}y=\frac{2L}{\pi D}$$ ■

例 2e(正态分布的刻画) 一颗子弹打在靶子上，令 X 和 Y 分别表示子弹的弹着点距靶心的水平偏差和垂直偏差，并且假定

1. X 和 Y 为独立的连续型随机变量，且具有可微密度函数.

251

2. X 和 Y 的联合密度 $f(x,y)=f_X(x)f_Y(y)$ 作为(x,y)的函数只依赖 x^2+y^2.

粗略地讲，假定(2)表示子弹射中 $x-y$ 平面内任意一点的概率，仅依赖于该点到靶心的距离而不依赖于其方向角. 假定(2)的另外一种等价描述为联合密度函数是旋转不变量.

更有趣的是，假定(1)与(2)隐含着 X 和 Y 都是正态分布的随机变量这一事实. 为了证明这一点，首先注意到由上面的假定可以导出关系式

$$f(x,y)=f_X(x)f_Y(y)=g(x^2+y^2) \tag{2-9}$$

其中 g 为某一个函数，将式(2-9)两边关于 x 求导，有

$$f'_X(x)f_Y(y)=2xg'(x^2+y^2) \tag{2-10}$$

用式(2-10)除以式(2-9)得

$$\frac{f'_X(x)}{f_X(x)}=\frac{2xg'(x^2+y^2)}{g(x^2+y^2)}$$

或

$$\frac{f'_X(x)}{2xf_X(x)}=\frac{g'(x^2+y^2)}{g(x^2+y^2)} \tag{2-11}$$

式(2-11)左边的值仅仅依赖于 x，而右边的值依赖于 x^2+y^2，由此可知左边对所有 x 是相同的. 实际上，考虑任意的 x_1，x_2，使得 y_1，y_2 满足 $x_1^2+y_1^2=x_2^2+y_2^2$. 那么，从式(2-11)可以得到

$$\frac{f'_X(x_1)}{2x_1f_X(x_1)}=\frac{g'(x_1^2+y_1^2)}{g(x_1^2+y_1^2)}=\frac{g'(x_2^2+y_2^2)}{g(x_2^2+y_2^2)}=\frac{f'_X(x_2)}{2x_2f_X(x_2)}$$

因此

$$\frac{f'_X(x)}{xf_X(x)}=c \text{ 或者 } \frac{\mathrm{d}}{\mathrm{d}x}(\log f_X(x))=cx$$

两边同时积分，则有

$$\log f_X(x)=a+\frac{cx^2}{2} \text{ 或者 } f_X(x)=k\mathrm{e}^{cx^2/2}$$

因为 $\int_{-\infty}^{\infty}f_X(x)\mathrm{d}x=1$，所以 c 一定是负值，可以写作 $c=-1/\sigma^2$. 因此，

$$f_X(x)=k\mathrm{e}^{-x^2/2\sigma^2}$$

也就是说，X 是一个正态随机变量，其中参数为 $\mu=0$ 和 σ^2. 同理可得

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\,\bar{\sigma}}\mathrm{e}^{-y^2/2\bar{\sigma}^2}$$

[252]

而且，由假定(2)可知 $\sigma^2=\bar{\sigma}^2$，则 X 和 Y 是服从参数 $\mu=0$ 和 σ^2 的独立同分布正态随机变量. ■

X 和 Y 独立的充分必要条件是它们的联合概率密度函数(或者是离散情形下的联合概率质量函数) $f(x,y)$ 可以分成两部分，一部分仅仅依赖于 x，另一部分仅仅依赖于 y.

命题 2.1 *连续(离散)型随机变量 X 和 Y 是独立的，当且仅当其联合概率密度(质量)函数可以表示为*

$$f_{X,Y}(x,y)=h(x)g(y) \qquad -\infty<x<\infty,-\infty<y<\infty$$

证明 下面给出连续情形下的证明. 首先注意独立性意味着联合密度是 X 和 Y 的边缘密度函数的乘积，所以当随机变量独立时，上述分解是成立的. 现在假设

$$f_{X,Y}(x,y)=h(x)g(y)$$

则

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{\infty}h(x)\mathrm{d}x\int_{-\infty}^{\infty}g(y)\mathrm{d}y=C_1C_2$$

其中 $C_1=\int_{-\infty}^{\infty}h(x)\mathrm{d}x$ 和 $C_2=\int_{-\infty}^{\infty}g(y)\mathrm{d}y$，且

$$f_X(x)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)\mathrm{d}y=C_2h(x)$$

$$f_Y(y)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)\mathrm{d}x=C_1g(y)$$

因为 $C_1C_2=1$，因此可以看到

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)$$

253 证毕. ■

例 2f 设 X 和 Y 的联合密度函数为

$$f(x,y)=\begin{cases}6\mathrm{e}^{-2x}\mathrm{e}^{-3y} & 0<x<\infty,0<y<\infty\\ 0 & \text{其他}\end{cases}$$

那么随机变量 X 和 Y 相互独立吗？设联合密度函数为

$$f(x,y)=\begin{cases}24xy & 0<x<1,0<y<1,0<x+y<1\\ 0 & \text{其他}\end{cases}$$

则随机变量相互独立吗？

解 对第一个例子，由于联合密度函数可进行因式分解，因此随机变量是相互独立的（其中一个是参数为 2 的指数分布，另外一个是参数为 3 的指数分布）. 对第二个例子，因为联合密度非 0 的区域不能用形式 $x\in A$，$y\in B$ 表示，联合密度不可进行因式分解，所以随机变量不独立. 可以用以下情形清楚地表示：

$$I(x,y)=\begin{cases}1 & \text{若 } 0<x<1,0<y<1,0<x+y<1\\ 0 & \text{其他}\end{cases}$$

可以写成

$$f(x,y)=24xyI(x,y)$$

显然这个式子不能分成两部分，一部分仅仅依赖于 x，另一部分仅仅依赖于 y. ■

独立性的概念可以推广到两个以上的随机变量. 一般来讲，n 个随机变量 X_1，X_2，…，X_n 称为相互独立的，如果对一切的实数集合 A_1，A_2，…，A_n，有

$$P\{X_1\in A_1,X_2\in A_2,\cdots,X_n\in A_n\}=\prod_{i=1}^{n}P\{X_i\in A_i\}$$

如前所证，这个条件等价于对一切 a_1，a_2，…，a_n，有

$$P\{X_1\leqslant a_1,X_2\leqslant a_2,\cdots,X_n\leqslant a_n\}=\prod_{i=1}^{n}P\{X_i\leqslant a_i\}$$

最后，我们可以称一族无穷多个随机变量是独立的，如果其任一有限子族是独立的.

例 2g（计算机如何选择随机子集） 大部分计算机可以产生或者模拟(0,1)上均匀随机变量的值，通过内置的子程序(可以达到高度近似)产生这种随机数. 因此，对计算机来说模拟一个示性(即伯努利)随机变量是很容易的. 假设 I 是示性变量，使得

254

$$P\{I=1\}=p=1-P\{I=0\}$$

计算机可以通过选择(0,1)上的均匀随机数 U 模拟 I，令

$$I=\begin{cases}1 & \text{若 } U<p\\ 0 & \text{若 } U\geqslant p\end{cases}$$

假定我们在数字 $1,2,\cdots,n$ 中用计算机来选择 $k(k\leqslant n)$，使得大小为 k 的 $\binom{n}{k}$ 个子集中的每一个等可能地被选到. 现在给出一种计算机解决这个问题的方法. 为了产生这样的一个子集，我们首先依次模拟 n 个示性变量 $I_1,I_2,\cdots,I_n$，其中恰有 k 个等于 1. 那些满足 $I_i=1$ 的 i 组成所要求的子集.

为了生成随机变量 $I_1,I_2,\cdots,I_n$，从模拟 n 个独立的(0,1)上的均匀随机变量 U_1，U_2，…，

U_n 开始. 定义

$$I_1=\begin{cases}1 & U_1<\dfrac{k}{n}\\ 0 & 其他\end{cases}$$

一旦 $I_1,I_2,\cdots,I_i$ 被定义，然后令

$$I_{i+1}=\begin{cases}1 & U_{i+1}<\dfrac{k-(I_1+\cdots+I_i)}{n-i}\\ 0 & 其他\end{cases}$$

换句话说，在第 $i+1$ 步，以概率为子集中剩余的位置数(即 $k-\sum_{j=1}^{i}I_j$)除以剩余的可能的数(即 $n-i$)让 $I_{i+1}=1$(从而将 $i+1$ 放入所要求的子集之中). 因此，$I_1,I_2,\cdots,I_n$ 的联合分布定义为

$$P\{I_1=1\}=\frac{k}{n}$$

$$P\{I_{i+1}=1\mid I_1,I_2,\cdots,I_i\}=\frac{k-\sum_{j=1}^{i}I_j}{n-i}\qquad 1<i<n$$

由归纳法可以证明上述结果对 $k+n$ 中大小为 k 的所有子集是等可能的. 当 $k+n=2$ 时(即 $k=1,n=1$)，显然成立. 假定当 $k+n\leqslant l$ 时结论成立，现在假设 $k+n=l+1$ 且考虑任意大小为 k 的子集(比如 $i_1\leqslant i_2\leqslant\cdots\leqslant i_k$)，并分两种情形进行讨论. 255

情形 1: $i_1=1$

$$\begin{aligned}&P\{I_1=I_{i_2}=\cdots=I_{i_k}=1,其他\ I_j=0\}\\ &=P\{I_1=1\}P\{I_{i_2}=\cdots=I_{i_k}=1,其他\ I_j=0\mid I_1=1\}\end{aligned}$$

现在假定 $I_1=1$，子集的剩余的元素的选择和从 $n-1$ 个元素 $2,3,\cdots,n$ 中选择大小为 $k-1$ 的子集一样. 因此，由归纳假设，条件概率等于选择大小为 $k-1$ 的子集的概率，即 $1\Big/\binom{n-1}{k-1}$. 因此

$$P\{I_1=I_{i_2}=\cdots=I_{i_k}=1,其他\ I_j=0\}=\frac{k}{n}\frac{1}{\binom{n-1}{k-1}}=\frac{1}{\binom{n}{k}}$$

情形 2: $i_1\neq1$

$$\begin{aligned}&P\{I_{i_1}=I_{i_2}=\cdots=I_{i_k}=1,其他\ I_j=0\}\\ &=P\{I_{i_1}=\cdots=I_{i_k}=1,其他\ I_j=0\mid I_1=0\}P\{I_1=0\}\\ &=\frac{1}{\binom{n-1}{k}}\left(1-\frac{k}{n}\right)=\frac{1}{\binom{n}{k}}\end{aligned}$$

其中利用归纳假设来计算上述条件概率.

因此，在所有的情形下，选择大小为 k 的子集的概率为 $1\Big/\binom{n}{k}$. ■

注 上述生成随机子集的方法占用计算机的内存很小. 一种快速算法需要更大的内存，参见 10.1 节. (后一种算法利用了 $1,2,\cdots,n$ 的随机排列的最后 k 个元素.)

例 2h 设 X,Y,Z 是独立的且服从$(0,1)$上的均匀分布. 计算 $P\{X\geqslant YZ\}$.

解 因为

$$f_{X,Y,Z}(x,y,z)=f_X(x)f_Y(y)f_Z(z)=1 \quad 0\leqslant x\leqslant 1, 0\leqslant y\leqslant 1, 0\leqslant z\leqslant 1$$

256

则有

$$P\{X\geqslant YZ\}=\iiint_{x\geqslant yz} f_{X,Y,Z}(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z=\int_0^1\int_0^1\int_{yz}^1 \mathrm{d}x\mathrm{d}y\mathrm{d}z$$

$$=\int_0^1\int_0^1(1-yz)\mathrm{d}y\mathrm{d}z=\int_0^1\left(1-\frac{z}{2}\right)\mathrm{d}z=\frac{3}{4}$$ ■

例 2i(半衰期的概率解释) 令 $N(t)$表示放射性物质在时刻 t 包含的原子核数. 半衰期的概念通常以某一确定的方式定义为：对给定的 h，存在一个经验式子，称为半衰期，即

$$N(t)=2^{-t/h}N(0) \qquad t>0$$

[注意 $N(h)=N(0)/2$.]因为上述式子隐含着对任意非负值 s,t，有

$$N(t+s)=2^{-(s+t)/h}N(0)=2^{-t/h}N(s)$$

所以不论已消逝多长时间 s，在剩下的时间 t 里，存在的原子核数将按照 $2^{-t/h}$ 递减.

上述确定性关系给出了从包含大量核子数的放射性物质的观察中所得到的结果，这与概率解释是一致的. 推导半衰期的合适概率模型的方法来自于经验观察，在任何时间间隔内衰变的比例既不依赖开始时间间隔内的总原子核数，也不依赖在这个时间间隔内的位置(因为 $N(t+s)/N(s)$ 既不依赖于 $N(s)$，也不依赖于 s). 因此，这表明各个原子核的行为是独立的，服从无记忆寿命分布. 因为唯一的无记忆分布是指数分布，同时因为给定的物质每 h 个时间单元衰变一半，因此我们提出下面的放射性衰减的概率模型.

半衰期 h 的概率解释：单个原子核的寿命是独立的随机变量，服从均值为 h 的指数分布. 也就是说，若 L 表示一个给定原子核的寿命，则

257

$$P\{L<t\}=1-2^{-t/h}$$

(因为 $P\{L<h\}=\frac{1}{2}$且上式可以写成

$$P\{L<t\}=1-\exp\left\{-t\frac{\log 2}{h}\right\}$$

可以看到 L 实际上是均值为 h 的指数分布.)

应该注意到，在前面半衰期的概率解释中，若在 0 时刻有 $N(0)$个原子核，则在时刻 t 时剩余的原子核数为 $N(t)$，$N(t)$是参数为 $n=N(0)$，$p=2^{-t/h}$ 的二项分布. 第 8 章的结果将证明，当在所给定的时间间隔内衰变的原子核的数目很大时，半衰期的这种解释和确定性模型是一致的. 然而，当考虑原子核的真实数量时，确定性关系和概率解释之间的差异是很明显的. 现在，我们将简要说明质子是否衰变的问题.

关于质子是否衰变有些争议. 事实上，有一种理论预测质子大约在 $h=10^{30}$ 年的半衰期内衰变. 从经验上来说，为了检验其是否正确，建议在一或两年观察大量质子，确定每一个质子在这段时间内是否衰变. (显然，在半衰期为 $h=10^{30}$ 年内观察大量质子是否衰变是不可行的.) 假定我们在 c 年能够观测 $N(0)=10^{30}$ 个质子. 根据确定性模型所预测的衰变量由下式给出：

$$N(0)-N(c)=h(1-2^{-c/h})=\frac{1-2^{-c/h}}{1/h}$$

$$\approx\lim_{x\to 0}\frac{1-2^{-cx}}{x} \qquad 因为\frac{1}{h}=10^{-30}\approx 0$$

$$=\lim_{x\to 0}(c2^{-cx}\log 2) \qquad 由洛必达法则$$

$$=c\log 2\approx 0.6931c$$

例如，两年内确定性模型估计将有 1.386 3 个质子衰变. 如果在这两年期间没有观察到衰变，将是对质子半衰期为 10^{30} 年这个假设的重要一击.

现在将上面的结论和从概率模型得到的结果进行比较. 再次考虑质子的半衰期为 $h=10^{30}$ 年这个假定，跟踪 h 个质子 c 年，因为存在大量相对独立的质子，每一个在这个时间间隔内将有很小的概率衰变，所以衰变的质子数量将服从参数为 $h(1-2^{-c/h})\approx c\log 2$ 的泊松分布. 因此 [258]

$$P(0\text{ 衰变})=e^{-c\log 2}=e^{-\log(2^c)}=\frac{1}{2^c}$$

一般而言，

$$P(n\text{ 衰变})=\frac{2^{-c}[c\log 2]^n}{n!}\qquad n\geqslant 0$$

于是我们看到即使两年内衰变的平均数(由确定性模型预测)为 1.386 3，在 4 个中也有一个不会发生任何衰变，因此表明这个结果绝不会使得最初的质子衰变假设无意义. ■

注 独立性是对称的关系. 随机变量 X 和 Y 是独立的，如果它们的联合密度函数(或离散情形下的质量函数)是它们各自密度(或质量)函数的乘积. 因此，称 X 与 Y 独立等价于称 Y 与 X 独立的，或仅说 X 与 Y 是独立的. 因此，在已知 Y 的值不会改变 X 的概率这一条件不直观的情况下，考虑 X 与 Y 的交换律比考虑 X 与 Y 独立更为有利，即考虑 Y 与 X 是否独立. 下面的例子说明了这一点.

例 2j 在掷骰子的过程中，最初骰子之和为 4，那么掷骰子的人继续，直到掷得的点数之和为 4 或者 7. 如果点数之和为 4，此人赢，如果点数之和为 7，此人输. 设 N 表示直到 4 或者 7 出现需要掷的次数，X 表示最后掷得的点数(4 或者 7). N 是否与 X 独立呢？也就是说，4 或者 7 首次出现是否影响这个数的出现所需掷的次数的分布. 大多数人从直觉上不能明显地找到这个问题的答案. 然而，反过来问 X 与 N 是否独立？也就是说，已知掷多少次骰子得到了点数之和为 4 或者 7，是否影响点数之和为 4 的概率？例如，设掷 n 次骰子可以得到点数之和为 4 或者 7，这影响最终和的概率分布吗？很明显不影响，因为最重要的是点数之和要么是 4 要么是 7，而在前 $n-1$ 次掷的过程中没有一次是 4 或 7，这个事实并不改变第 n 次掷的概率. 因此可以推出 X 独立于 N 的结论，即 N 与 X 是独立的.

再举一个例子，设 $X_1,X_2,\cdots$ 是独立同分布的连续随机变量序列，且依次观察这些随机变量. 如果对每个 $i=1,\cdots,n-1$，$X_n>X_i$，则称 X_n 是记录值(record value). 也就是说，比前面大的随机变量称为记录值. 令 A_n 表示 X_n 是记录值的事件，A_{n+1} 与 A_n 是否独立？也就是说，已知第 n 个随机变量是前 n 个随机变量的最大值是否影响第 $n+1$ 个随机变量是前 $n+1$ 个随机变量的最大值的概率？如果这种情况成立，则 A_{n+1} 独立于 A_n，但这在直观上并不明显. 然而，反过来考虑 A_n 是否独立于 A_{n+1}，那么更加容易理解. 对于已知第 $n+1$ 个变量大于 $X_1,X_2,\cdots,X_n$，很明显提供不了任何关于前 n 个随机变量中 X_n 的相对大小的信息. 事实上，由对称性很容易知道，n 个随机变量中的任何一个都等可能地是这个集中最大的一个. 即 $P(A_n|A_{n+1})=P(A_n)=1/n$，因此推出 A_n 与 A_{n+1} 是相互独立的事件的结论. ■ [259]

注 从等式

$$\begin{aligned}&P\{X_1\leqslant a_1,\cdots,X_n\leqslant a_n\}\\&=P\{X_1\leqslant a_1\}P\{X_2\leqslant a_2|X_1\leqslant a_1\}\cdots P\{X_n\leqslant a_n|X_1\leqslant a_1,\cdots,X_{n-1}\leqslant a_{n-1}\}\end{aligned}$$

可以分别建立 $X_1, X_2, \cdots, X_n$ 的独立性. 即，要证明随机变量相互独立，只要证明

X_2 独立于 X_1

X_3 独立于 X_1, X_2

X_4 独立于 X_1, X_2, X_3

……

X_n 独立于 $X_1, \cdots, X_{n-1}$

6.3 独立随机变量之和

当 X 和 Y 独立时，从 X,Y 的分布来计算 $X+Y$ 的分布是非常重要的. 假定 X,Y 是独立连续型随机变量，概率密度函数分别为 f_X, f_Y，则 $X+Y$ 的累积分布函数可以如下计算得到：

$$\begin{aligned} F_{X+Y}(a) &= P\{X+Y \leqslant a\} = \iint\limits_{x+y\leqslant a} f_X(x) f_Y(y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x) f_Y(y)\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)\mathrm{d}x f_Y(y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\mathrm{d}y \end{aligned} \tag{3-1}$$

260

累积分布函数 F_{X+Y} 称为 F_X 与 F_Y 的卷积（F_X, F_Y 分别为 X,Y 的累积分布函数).

对式(3-1)两边求导，可以得到 $X+Y$ 的概率密度函数 f_{X+Y} 为

$$\begin{aligned} f_{X+Y}(a) &= \frac{\mathrm{d}}{\mathrm{d}a}\int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\mathrm{d}y = \int_{-\infty}^{\infty} \frac{\mathrm{d}}{\mathrm{d}a} F_X(a-y) f_Y(y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)\mathrm{d}y \end{aligned} \tag{3-2}$$

例 3a（两个独立的均匀分布随机变量之和） 设 X 和 Y 是独立随机变量，服从(0,1)上的均匀分布，计算 $X+Y$ 的概率密度.

解 根据式(3-2)，因为

$$f_X(a) = f_Y(a) = \begin{cases} 1 & 0<a<1 \\ 0 & \text{其他} \end{cases}$$

可得

$$f_{X+Y}(a) = \int_0^1 f_X(a-y)\mathrm{d}y$$

对 $0 \leqslant a \leqslant 1$，有

$$f_{X+Y}(a) = \int_0^a \mathrm{d}y = a$$

对 $1<a<2$，有

$$f_{X+Y}(a) = \int_{a-1}^1 \mathrm{d}y = 2-a$$

因此

$$f_{X+Y}(a) = \begin{cases} a & 0 \leqslant a \leqslant 1 \\ 2-a & 1<a<2 \\ 0 & \text{其他} \end{cases}$$

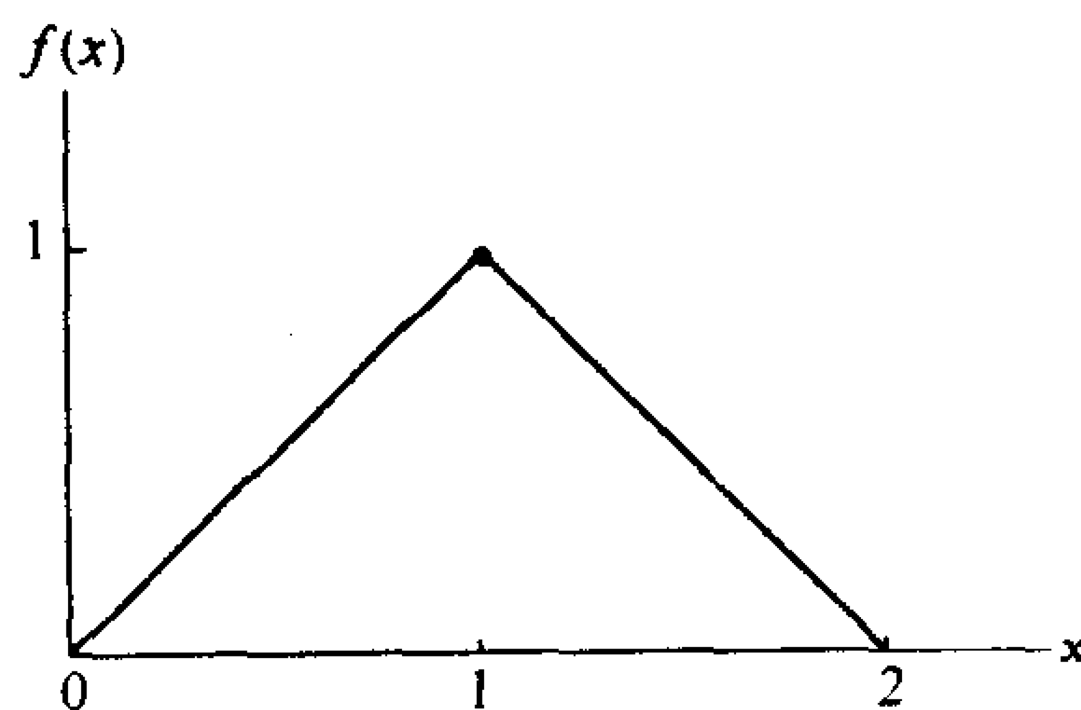


图 6-3 三角密度函数

由于密度函数的形状为三角形(见图 6-3)，所以称随机变量 $X+Y$ 服从三角分布(triangular distribution). ■

261

回忆 Γ 随机变量的密度函数为

$$f(y)=\frac{\lambda e^{-\lambda y}(\lambda y)^{t-1}}{\Gamma(t)} \qquad 0<y<\infty$$

这个分布函数族的一个重要性质为对于固定的 λ 值，卷积封闭.

命题 3.1 设 X 和 Y 分别是参数为 (s,λ) 和 (t,λ) 的独立 Γ 随机变量，则 $X+Y$ 也是 Γ 随机变量，其参数为 $(s+t,\lambda)$.

证明 根据式(3-2)，可得

$$\begin{aligned} f_{X+Y}(a) &= \frac{1}{\Gamma(s)\Gamma(t)}\int_0^a \lambda e^{-\lambda(a-y)}[\lambda(a-y)]^{s-1}\lambda e^{-\lambda y}(\lambda y)^{t-1}dy \\ &= Ke^{-\lambda a}\int_0^a (a-y)^{s-1}y^{t-1}dy \\ &= Ke^{-\lambda a}a^{s+t-1}\int_0^1 (1-x)^{s-1}x^{t-1}dx \qquad 令\ x=\frac{y}{a} \\ &= Ce^{-\lambda a}a^{s+t-1} \end{aligned}$$

其中 C 是不依赖于 a 的常量. 因为是密度函数，所以积分为 1，可以确定 C 的值，有

$$f_{X+Y}(a)=\frac{\lambda e^{-\lambda a}(\lambda a)^{s+t-1}}{\Gamma(s+t)}$$

则结论得证. ■

根据命题 3.1 和归纳法可以容易地证明：设 $X_i(i=1,\cdots,n)$ 是独立的 Γ 随机变量，参数分别为 $(t_i,\lambda),i=1,\cdots,n$，则 $\sum_{i=1}^n X_i$ 也是 Γ 随机变量，参数为 $\left(\sum_{i=1}^n t_i,\lambda\right)$. 证明留作练习.

例 3b 设 $X_1,X_2,\cdots X_n$ 是 n 个参数为 λ 的独立的指数随机变量，则因为参数为 λ 的指数随机变量也就是参数为 $(1,\lambda)$ 的 Γ 随机变量，所以根据命题 3.1 可得 $X_1+X_2+\cdots+X_n$ 为参数为 (n,λ) 的 Γ 随机变量.

262

如果 $Z_1,Z_2,\cdots,Z_n$ 是独立的标准正态随机变量，则 $Y\equiv\sum_{i=1}^n Z_i^2$ 是服从自由度为 n 的卡方(chi-square，有时写作 χ^2)分布. 我们来计算概率密度函数. 当 $n=1,Y=Z_1^2$ 时，从第 5 章例 7b 可得概率密度函数为

$$f_{Z^2}(y)=\frac{1}{2\sqrt{y}}[f_Z(\sqrt{y})+f_Z(-\sqrt{y})]=\frac{1}{2\sqrt{y}}\frac{2}{\sqrt{2\pi}}e^{-\frac{y}{2}}=\frac{\frac{1}{2}e^{-y/2}(y/2)^{\frac{1}{2}-1}}{\sqrt{\pi}}$$

上式是参数为 $\left(\frac{1}{2},\frac{1}{2}\right)$ 的 Γ 分布. $\left[这一分析的一个副产物是 \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}.\right]$ 如果每一个 Z_i^2 服从 $\Gamma\left(\frac{1}{2},\frac{1}{2}\right)$，则由命题 3.1 知自由度为 n 的 χ^2 分布就是参数为 $\left(\frac{n}{2},\frac{1}{2}\right)$ 的 Γ 分布，因此概率密度函数为

$$\begin{aligned} f_{\chi^2}(y) &= \frac{\frac{1}{2}e^{-y/2}\left(\frac{y}{2}\right)^{n/2-1}}{\Gamma\left(\frac{n}{2}\right)} \qquad y>0 \\ &= \frac{e^{-y/2}y^{n/2-1}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \qquad y>0 \end{aligned}$$

当 n 是偶数时，$\Gamma(n/2)=[(n/2)-1]!$，而当 n 是奇数时，$\Gamma(n/2)$ 可以由迭代关系式 $\Gamma(t)=(t-1)\Gamma(t-1)$，

并且根据$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$而得到. $\left[\text{例如}, \Gamma\left(\frac{5}{2}\right)=\frac{3}{2}\Gamma\left(\frac{3}{2}\right)=\frac{3}{2}\frac{1}{2}\Gamma\left(\frac{1}{2}\right)=\frac{3}{4}\sqrt{\pi}.\right]$

在实际中卡方分布通常作为均方误差的分布而出现，此误差是试图打中 n 维空间中的某一目标时产生的，并且认为坐标误差是独立的标准正态随机变量. 卡方分布在统计分析中也很重要.

263

用式(3-2)也可以证明下面关于正态随机变量的重要结论.

命题 3.2 设 $X_i(i=1,\cdots,n)$是服从正态分布的独立随机变量，其参数分别为 $\mu_i,\sigma_i^2,i=1,\cdots,n$，则 $\sum_{i=1}^n X_i$ 也服从正态分布，其参数为 $\sum_{i=1}^n \mu_i$ 和 $\sum_{i=1}^n \sigma_i^2$.

证明 首先，设 X 和 Y 为独立的正态随机变量，其中 X 的均值为 0，方差为 σ^2；Y 的均值为 0，方差为 1. 我们利用式(3-2)来确定 $X+Y$ 的密度函数. 现在，根据

$$c=\frac{1}{2\sigma^2}+\frac{1}{2}=\frac{1+\sigma^2}{2\sigma^2}$$

我们有

$$\begin{aligned}f_X(a-y)f_Y(y)&=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(a-y)^2}{2\sigma^2}\right\}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\}\\&=\frac{1}{2\pi\sigma}\exp\left\{-\frac{a^2}{2\sigma^2}\right\}\exp\left\{-c\left(y^2-2y\,\frac{a}{1+\sigma^2}\right)\right\}\end{aligned}$$

因此，由式(3-2)，

$$\begin{aligned}f_{X+Y}(a)&=\frac{1}{2\pi\sigma}\exp\left\{-\frac{a^2}{2\sigma^2}\right\}\exp\left\{\frac{a^2}{2\sigma^2(1+\sigma^2)}\right\}\int_{-\infty}^{\infty}\exp\left\{-c\left(y-\frac{a}{1+\sigma^2}\right)^2\right\}\mathrm{d}y\\&=\frac{1}{2\pi\sigma}\exp\left\{-\frac{a^2}{2(1+\sigma^2)}\right\}\int_{-\infty}^{\infty}\exp\{-cx^2\}\mathrm{d}x=C\exp\left\{-\frac{a^2}{2(1+\sigma^2)}\right\}\end{aligned}$$ ■

这里 C 与 a 无关. 但结果表明 $X+Y$ 是正态的且均值为 0、方差为 $1+\sigma^2$.

现在，假设 X_1 和 X_2 是独立的正态随机变量，其中 X_i 的均值和方差分别为μ_i，$\sigma_i^2,i=1,2$. 那么

$$X_1+X_2=\sigma_2\left(\frac{X_1-\mu_1}{\sigma_2}+\frac{X_2-\mu_2}{\sigma_2}\right)+\mu_1+\mu_2$$

但是因为$(X_1-\mu_1)/\sigma_2$ 服从均值为 0、方差为 σ_1^2/σ_2^2 的正态分布；$(X_2-\mu_2)/\sigma_2$ 服从均值为 0、方差为 1 的正态分布，因此由上面的结论得出$(X_1-\mu_1)/\sigma_2+(X_2-\mu_2)/\sigma_2$ 服从均值为 0、方差为 $1+\sigma_1^2/\sigma_2^2$ 的正态分布，从而表明 X_1+X_2 服从均值为 $\mu_1+\mu_2$、方差为 $\sigma_2^2(1+\sigma_1^2/\sigma_2^2)=\sigma_1^2+\sigma_2^2$ 的正态分布.

264

因而，当 $n=2$ 时命题 3.2 成立. 现在利用归纳法推导出一般情形. 假设有 $n-1$ 个随机变量时，命题成立. 现在考虑有 n 个随机变量的情形，记

$$\sum_{i=1}^n X_i=\sum_{i=1}^{n-1} X_i+X_n$$

通过归纳假设，得 $\sum_{i=1}^{n-1} X_i$ 服从均值为 $\sum_{i=1}^{n-1}\mu_i$ 、方差为 $\sum_{i=1}^{n-1}\sigma_i^2$ 的正态分布 . 因此，根据 $n=2$ 时的结论，可以得出 $\sum_{i=1}^{n} X_i$ 服从均值为 $\sum_{i=1}^{n}\mu_i$ 、方差为 $\sum_{i=1}^{n}\sigma_i^2$ 的正态分布.

例 3c 某俱乐部的篮球队将某一赛季举行 44 场比赛. 其中有 26 场比赛要和 A 级队进

行，18 场比赛要和 B 级队进行．假如这个球队每场比赛打赢 A 级队的概率为 0.4，每场比赛打赢 B 级队的概率为 0.7. 同时假设不同比赛的结果是相互独立的，估计以下概率：

(a) 球队至少打赢 25 场比赛.

(b) 球队打赢 A 级队的比赛要比打赢 B 级队的比赛多．

解 (a) 记 X_A，X_B 分别表示打赢 A 级队和打赢 B 级队的比赛场数，并且 X_A 和 X_B 是相互独立的二项随机变量，则有

$$E[X_A]=26\times(0.4)=10.4 \qquad \text{Var}(X_A)=26\times(0.4)\times(0.6)=6.24$$
$$E[X_B]=18\times(0.7)=12.6 \qquad \text{Var}(X_B)=18\times(0.7)\times(0.3)=3.78$$

根据二项分布的正态逼近推导出 X_A 和 X_B 相互独立、将近似服从具有前面给定的期望值和方差的正态分布．因此，根据命题 3.2，X_A+X_B 将近似服从均值为 23、方差为 10.02 的正态分布. 因而，以 Z 来表示一个标准正态随机变量，我们有

$$P\{X_A+X_B\geqslant 25\}=P\{X_A+X_B\geqslant 24.5\}=P\left\{\frac{X_A+X_B-23}{\sqrt{10.02}}\geqslant\frac{24.5-23}{\sqrt{10.02}}\right\}$$
$$\approx P\left\{Z\geqslant\frac{1.5}{\sqrt{10.02}}\right\}\approx 1-P\{Z<0.4739\}\approx 0.3178$$

265

(b) 注意到 X_A-X_B 将近似服从均值为 -2.2、方差为 10.02 的正态分布．因而有

$$P\{X_A-X_B\geqslant 1\}=P\{X_A-X_B\geqslant 0.5\}=P\left\{\frac{X_A-X_B+2.2}{\sqrt{10.02}}\geqslant\frac{0.5+2.2}{\sqrt{10.02}}\right\}$$
$$\approx P\left\{Z\geqslant\frac{2.7}{\sqrt{10.02}}\right\}\approx 1-P\{Z<0.8530\}\approx 0.1968$$

因此，球队至少打赢 25 场比赛的机会约占 31.78%，打赢 A 级队要比打赢 B 级队的比赛场数多的机会约占 19.68%. ■

如果 $\log(Y)$ 是均值为 μ、方差为 σ^2 的正态随机变量，则随机变量 Y 被称为参数为 μ 和 σ 的对数正态随机变量. 即 Y 是对数正态的，如果它可表示为

$$Y=e^X$$

其中，X 是正态随机变量.

例 3d 从某一固定的时间开始，令 $S(n)$ 表示某种证券在 $n(n\geqslant 1)$ 个星期后的价格. 对于这些价格的一般模型，假定其价格比为 $S(n)/S(n-1)(n\geqslant 1)$ 是独立同分布的对数正态随机变量. 假设此模型的参数为 $\mu=0.0165$，$\sigma=0.0730$，求以下情形的概率：

(a) 每隔两个星期证券价格都增长.

(b) 两星期后证券价格比现在高.

解 记 Z 为一个标准正态随机变量. 为解(a)，我们利用函数 $\log(x)$ 随 x 的增加而增加的性质得出：$x>1$ 当且仅当 $\log(x)>\log(1)=0$. 因此，我们有

$$P\left\{\frac{S(1)}{S(0)}>1\right\}=P\left\{\log\left(\frac{S(1)}{S(0)}\right)>0\right\}=P\left\{Z>\frac{-0.0165}{0.0730}\right\}=P\{Z<0.2260\}=0.5894$$

因此，一星期后价格上涨的概率为 0.5894. 由于连续价格比是相互独立的，所以每隔两星期证券价格增长的概率是 $(0.5894)^2=0.3474$.

266

为解(b)，推理如下：

$$P\left\{\frac{S(2)}{S(0)}>1\right\}=P\left\{\frac{S(2)}{S(1)}\,\frac{S(1)}{S(0)}>1\right\}=P\left\{\log\left(\frac{S(2)}{S(1)}\right)+\log\left(\frac{S(1)}{S(0)}\right)>0\right\}$$

但是，$\log\left(\frac{S(2)}{S(1)}\right)+\log\left(\frac{S(1)}{S(0)}\right)$是两个相互独立的正态随机变量的和，且两个随机变量有共同的均值 0.016 5 和共同的标准差 0.073 0，则两者的和是一个均值为 0.033 0、方差为 $2(0.0730)^2$ 的正态随机变量.

从而，

$$P\left\{\frac{S(2)}{S(0)}>1\right\}=P\left\{Z>\frac{-0.0330}{0.0730\sqrt{2}}\right\}=P\{Z<0.31965\}=0.6254$$ ■

在离散情况下，我们将考虑以下例子，而不是试图得到 $X+Y$ 的分布的一般表达式.

例 3c（独立泊松随机变量之和） 设 X,Y 分别为以 λ_1,λ_2 为参数的独立泊松随机变量，求 $X+Y$ 的分布.

解 因为事件$\{X+Y=n\}$可以写成互不相交的事件$\{X=k,Y=n-k\}(0\leqslant k\leqslant n)$的并，则

$$\begin{aligned}P\{X+Y=n\}&=\sum_{k=0}^{n}P\{X=k,Y=n-k\}=\sum_{k=0}^{n}P\{X=k\}P\{Y=n-k\}\\&=\sum_{k=0}^{n}e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}=e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!}\\&=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n\end{aligned}$$

就是说，$X+Y$ 是以 $\lambda_1+\lambda_2$ 为参数的泊松分布. ■

例 3f（独立二项随机变量之和） 设 X,Y 分别为以$(n,p),(m,p)$为参数的独立二项随机变量，求 $X+Y$ 的分布.

267

解 回忆对二项随机变量的解释，不需任何计算知，$X+Y$ 是以$(n+m,p)$为参数的二项随机变量. 这是因为 X 表示 n 次独立重复试验中成功的次数，且每次试验成功的概率为 p；类似地，Y 表示 m 次独立重复试验中成功的次数，且每次试验成功的概率为 p. 因此，若 X 与 Y 是独立的，那么 $X+Y$ 表示 $n+m$ 次独立试验中成功的次数，且每次试验成功的概率为 p. 因此，$X+Y$ 是以$(n+m,p)$为参数的二项随机变量. 为了验证这一结果，注意

$$\begin{aligned}P\{X+Y=k\}&=\sum_{i=0}^{n}P\{X=i,Y=k-i\}=\sum_{i=0}^{n}P\{X=i\}P\{Y=k-i\}\\&=\sum_{i=0}^{n}\binom{n}{i}p^iq^{n-i}\binom{m}{k-i}p^{k-i}q^{m-k+i}\end{aligned}$$

其中 $q=1-p$，当 $j>r$ 时，$\binom{r}{j}=0$. 故

$$P\{X+Y=k\}=p^kq^{n+m-k}\sum_{i=0}^{n}\binom{n}{i}\binom{m}{k-i}$$

应用下面的组合等式即得结论：

$$\binom{n+m}{k}=\sum_{i=0}^{n}\binom{n}{i}\binom{m}{k-i}$$ ■

6.4 条件分布：离散情形

回忆对任意两个事件 E 和 F，当 $P(F)>0$ 时，已知 F 发生的条件下 E 的条件概率为

$$P(E|F)=\frac{P(EF)}{P(F)}$$

因此，如果 X 和 Y 是离散型随机变量，那么对一切使得 $p_Y(y)>0$ 的 y，我们很自然地把已知 $Y=y$ 的条件下 X 的条件概率质量函数定义为

$$p_{X|Y}(x|y)=P\{X=x|Y=y\}=\frac{P\{X=x,Y=y\}}{P\{Y=y\}}=\frac{p(x,y)}{p_Y(y)}$$

268

同理，对一切使得 $p_Y(y)>0$ 的 y，我们定义已知 $Y=y$ 的条件下 X 的条件概率分布函数为

$$F_{X|Y}(x \mid y) = P\{X \leqslant x \mid Y = y\} = \sum_{a\leqslant x} p_{X|Y}(a \mid y)$$

换句话说，除了把各项取成在事件 $Y=y$ 的条件下以外，上述定义与无条件的情形完全一样. 如果 X 与 Y 独立，则 X 对于 Y 的条件质量函数以及条件分布函数与无条件的情形相同. 这是因为，如果 X 与 Y 独立，那么

$$p_{X|Y}(x|y)=P\{X=x|Y=y\}=\frac{P\{X=x,Y=y\}}{P\{Y=y\}}=\frac{P\{X=x\}P\{Y=y\}}{P\{Y=y\}}=P\{X=x\}$$

例 4a 设 X 和 Y 的联合概率质量函数 $p(x,y)$ 为

$$p(0,0)=0.4 \quad p(0,1)=0.2 \quad p(1,0)=0.1 \quad p(1,1)=0.3$$

求已知 $Y=1$ 的条件下 X 的条件概率质量函数.

解 我们首先指出

$$p_Y(1) = \sum_x p(x,1) = p(0,1)+p(1,1) = 0.5$$

因此

$$p_{X|Y}(0|1)=\frac{p(0,1)}{p_Y(1)}=\frac{2}{5}$$

$$p_{X|Y}(1|1)=\frac{p(1,1)}{p_Y(1)}=\frac{3}{5}$$

■

例 4b 设 X 和 Y 分别是以 λ_1 和 λ_2 为参数的独立泊松分布随机变量，求已知 $X+Y=n$ 的条件下 X 的条件分布函数.

269

解 计算 X 对于 $X+Y=n$ 的条件概率质量函数如下：

$$P\{X=k|X+Y=n\}=\frac{P\{X=k,X+Y=n\}}{P\{X+Y=n\}}=\frac{P\{X=k,Y=n-k\}}{P\{X+Y=n\}}=\frac{P\{X=k\}P\{Y=n-k\}}{P\{X+Y=n\}}$$

最后一个等式成立是根据 X 和 Y 的独立性假设而得. 回想(例 3d) $X+Y$ 具有以 $\lambda_1+\lambda_2$ 为参数的泊松分布，可见上式等于

$$P\{X=k|X+Y=n\}=\frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\left[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right]^{-1}=\frac{n!}{(n-k)!\ k!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n}$$

$$=\binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

换句话说，在已知 $X+Y=n$ 的条件下 X 的条件分布是以 n 以及 $\lambda_1/(\lambda_1+\lambda_2)$ 为参数的二项分布. ■

6.5 条件分布：连续情形

设 X 和 Y 的联合概率密度函数为 $f(x,y)$，则对于一切使 $f_Y(y)>0$ 的 y，在已知 $Y=y$ 的条件下 X 的条件概率密度函数为

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

为了阐明这个定义，将上式左边乘以 $\mathrm{d}x$，右边乘以 $(\mathrm{d}x\mathrm{d}y)/\mathrm{d}y$ 得

$$\begin{aligned}f_{X|Y}(x|y)\mathrm{d}x&=\frac{f(x,y)\mathrm{d}x\mathrm{d}y}{f_Y(y)\mathrm{d}y}\approx\frac{P\{x\leqslant X\leqslant x+\mathrm{d}x,y\leqslant Y\leqslant y+\mathrm{d}y\}}{P\{y\leqslant Y\leqslant y+\mathrm{d}y\}}\\&=P\{x\leqslant X\leqslant x+\mathrm{d}x|y\leqslant Y\leqslant y+\mathrm{d}y\}\end{aligned}$$

换句话说，对很小的 $\mathrm{d}x$，$\mathrm{d}y$，$f_{X|Y}(x|y)$ 表示已知 Y 在 y 与 $y+\mathrm{d}y$ 之间取值的条件下，X 在 x 与 $x+\mathrm{d}x$ 之间取值的条件概率.

270

运用条件概率密度函数，使我们在已知某一随机变量的值的条件下，可以定义与另一个随机变量有关的事件的条件概率. 也就是说，如果 X 和 Y 为联合连续型的，则对任意集 A，

$$P\{X\in A\mid Y=y\}=\int_A f_{X|Y}(x\mid y)\mathrm{d}x$$

特别地，取 $A=(-\infty,a]$，可以定义在已知 $Y=y$ 的条件下，X 的条件累积分布函数为

$$F_{X|Y}(a\mid y)\equiv P\{X\leqslant a\mid Y=y\}=\int_{-\infty}^{a}f_{X|Y}(x\mid y)\mathrm{d}x$$

读者应该注意，利用前面讨论中介绍的概念，即使作为条件的事件(即事件 $\{Y=y\}$)的概率等于 0，我们给出的条件概率的表达式仍然是适用的.

例 5a 设 X 和 Y 的联合密度为

$$f(x,y)=\begin{cases}\frac{15}{2}x(2-x-y) & 0<x<1,0<y<1\\0 & \text{其他}\end{cases}$$

求 X 在已知 $Y=y$ 条件下的条件密度，其中 $0<y<1$.

解 对 $0<x<1,0<y<1$，有

$$\begin{aligned}f_{X|Y}(x\mid y)&=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\int_{-\infty}^{\infty}f(x,y)\mathrm{d}x}=\frac{x(2-x-y)}{\int_0^1 x(2-x-y)\mathrm{d}x}\\&=\frac{x(2-x-y)}{\frac{2}{3}-\frac{y}{2}}=\frac{6x(2-x-y)}{4-3y}\end{aligned}$$

■

例 5b 设 X 和 Y 的联合密度为

$$f(x,y)=\begin{cases}\frac{\mathrm{e}^{-x/y}\mathrm{e}^{-y}}{y} & 0<x<\infty,0<y<\infty\\0 & \text{其他}\end{cases}$$

271

求 $P\{X>1|Y=y\}$.

解 先求 X 在已知 $Y=y$ 条件下的条件密度：

$$f_{X|Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)}=\frac{\mathrm{e}^{-x/y}\mathrm{e}^{-y}/y}{\mathrm{e}^{-y}\int_0^{\infty}(1/y)\mathrm{e}^{-x/y}\mathrm{d}x}=\frac{1}{y}\mathrm{e}^{-x/y}$$

因此，

$$P\{X>1\mid Y=y\}=\int_1^{\infty}\frac{1}{y}\mathrm{e}^{-x/y}\mathrm{d}x=-\mathrm{e}^{-x/y}\Big|_1^{\infty}=\mathrm{e}^{-1/y}$$

■

若 X 和 Y 是独立的连续型随机变量，则在已知 $Y=y$ 的条件下，X 的条件密度就是 X 的

无条件密度. 这是因为，在独立情形下，

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{f_X(x)f_Y(y)}{f_Y(y)}=f_X(x)$$

还可以讨论随机变量既不是联合连续也不是联合离散情况下的条件分布. 比如，X 是连续型随机变量，概率密度函数为 f，N 是离散型随机变量. 考虑给定 $N=n$ 时，X 的条件密度函数. 那么

$$\frac{P\{x<X<x+\mathrm{d}x|N=n\}}{\mathrm{d}x}=\frac{P\{N=n|x<X<x+\mathrm{d}x\}}{P\{N=n\}}\frac{P\{x<X<x+\mathrm{d}x\}}{\mathrm{d}x}$$

然后令 $\mathrm{d}x\to 0$，有

$$\lim_{\mathrm{d}x\to 0}\frac{P\{x<X<x+\mathrm{d}x|N=n\}}{\mathrm{d}x}=\frac{P\{N=n|X=x\}}{P\{N=n\}}f(x)$$

因此，在给定 $N=n$ 以后，X 的条件密度函数为

$$f_{X|N}(x|n)=\frac{P\{N=n|X=x\}}{P\{N=n\}}f(x)$$

例 5c 考虑具有共同的成功概率的 $n+m$ 次试验. 假如成功的概率事先不确定，但是从 (0,1) 的均匀分布中取值，求给定 $n+m$ 次试验中有 n 次成功的条件下成功概率的条件分布. 272

解 令 X 表示试验成功的概率，则 X 是 (0,1) 上均匀分布的随机变量. 另外，给定 $X=x$，则 $n+m$ 次试验是独立的且具有共同的成功概率 x，且设 N 表示成功的次数，它是服从参数为 $(n+m,x)$ 的二项随机变量. 因此给定 $N=n$ 的情况下，X 的条件密度为

$$\begin{aligned}f_{X|N}(x|n)&=\frac{P\{N=n|X=x\}f_X(x)}{P\{N=n\}}\\&=\frac{\binom{n+m}{n}x^n(1-x)^m}{P\{N=n\}}\qquad 0<x<1\\&=cx^n(1-x)^m\end{aligned}$$

其中 c 不依赖于 X. 因此，条件密度就是参数为 $n+1,m+1$ 的 β 随机变量.

前面的结果是非常有意义的，因为它表明如果试验成功的概率的先验分布（对收集的数据而言）服从 (0, 1) 上的均匀分布（或者等价地，服从参数为 (1,1) 的 β 分布），那么在 $n+m$ 次试验中有 n 次成功的后验分布（或者是条件分布）服从参数为 $(1+n,1+m)$ 的 β 分布. 这是非常有价值的，因为它加强了我们假设一个随机变量服从 β 分布的直觉. ■

*6.6 顺序统计量

设 $X_1,X_2,\cdots,X_n$ 为 n 个独立同分布的连续型随机变量，它们有共同的密度函数 f 和分布函数 F. 定义

$$\begin{aligned}X_{(1)}&=X_1,X_2,\cdots,X_n\text{ 中最小的}\\X_{(2)}&=X_1,X_2,\cdots,X_n\text{ 中第二小的}\\&\vdots\\X_{(j)}&=X_1,X_2,\cdots,X_n\text{ 中第 }j\text{ 小的}\\&\vdots\\X_{(n)}&=X_1,X_2,\cdots,X_n\text{ 中最大的}\end{aligned}$$

有序值 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 称为是对应于随机变量 $X_1,X_2,\cdots,X_n$ 的**顺序统计量**（order statistic）. 换句话说，$X_{(1)},X_{(2)},\cdots,X_{(n)}$ 是 $X_1,X_2,\cdots,X_n$ 的有序值.

顺序统计量 $X_{(1)},X_{(2)},\cdots,X_{(n)}$ 取值为 $x_1\leqslant x_2\leqslant\cdots\leqslant x_n$ 当且仅当对$(1,2,\cdots,n)$的某一排列$(i_1,i_2,\cdots,i_n)$

273

$$X_1=x_{i_1},X_2=x_{i_2},\cdots,X_n=x_{i_n}$$

据此可求出顺序统计量的联合密度函数. 因为对$(1,2,\cdots,n)$的任一排列$(i_1,i_2,\cdots,i_n)$,

$$P\left\{x_{i_1}-\frac{\varepsilon}{2}<X_1<x_{i_1}+\frac{\varepsilon}{2},\cdots,x_{i_n}-\frac{\varepsilon}{2}<X_n<x_{i_n}+\frac{\varepsilon}{2}\right\}$$

$$\approx\varepsilon^n f_{X_1,\cdots,X_n}(x_{i_1},\cdots,x_{i_n})=\varepsilon^n f(x_{i_1})\cdots f(x_{i_n})=\varepsilon^n f(x_1)\cdots f(x_n)$$

可以看到对 $x_1<x_2<\cdots<x_n$,

$$P\left\{x_1-\frac{\varepsilon}{2}<X_{(1)}<x_1+\frac{\varepsilon}{2},\cdots,x_n-\frac{\varepsilon}{2}<X_{(n)}<x_n+\frac{\varepsilon}{2}\right\}\approx n!\ \varepsilon^n f(x_1)\cdots f(x_n)$$

两边同除以 ε^n 并令 $\varepsilon\to0$ 得

$$f_{X_{(1)},\cdots,X_{(n)}}(x_1,x_2,\cdots,x_n)=n!\ f(x_1)\cdots f(x_n)\quad x_1<x_2<\cdots<x_n \tag{6-1}$$

式(6-1)的最简单解释如下：为使向量$\langle X_{(1)},\cdots,X_{(n)}\rangle$等于$\langle x_1,\cdots,x_n\rangle$，其充要条件为$\langle X_1,\cdots,X_n\rangle$等于$\langle x_1,\cdots,x_n\rangle$的 $n!$ 种排列之一. 因为$\langle X_1,\cdots,X_n\rangle$等于$\langle x_1,\cdots,x_n\rangle$的任一特定排列的概率(密度)恰为 $f(x_1)\cdots f(x_n)$，故得式(6-1).

例 6a 设有 3 人“随机地分布”在一条 1 英里长的公路上，求其中没有 2 人相距小于 d 英里的概率，这里 $d\leqslant\frac{1}{2}$.

解 假定“随机地分布”是指 3 人的位置相互独立而且在公路上有均匀分布. 令 X_i 表示第 i 个人的位置，则所求的概率为 $P\{X_{(i)}>X_{(i-1)}+d,i=2,3\}$. 因为

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x_1,x_2,x_3)=3!\qquad 0<x_1<x_2<x_3<1$$

所以

$$\begin{aligned}P\{X_{(i)}>X_{(i-1)}+d,i=2,3\}&=\iiint\limits_{\substack{x_i>x_{i-1}+d\\ i=2,3}}f_{X_{(1)},X_{(2)},X_{(3)}}(x_1,x_2,x_3)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\\&=3!\int_0^{1-2d}\int_{x_1+d}^{1-d}\int_{x_2+d}^{1}\mathrm{d}x_3\mathrm{d}x_2\mathrm{d}x_1=6\int_0^{1-2d}\int_{x_1+d}^{1-d}(1-d-x_2)\mathrm{d}x_2\mathrm{d}x_1\\&=6\int_0^{1-2d}\int_0^{1-2d-x_1}y_2\mathrm{d}y_2\mathrm{d}x_1\end{aligned}$$

274

其中我们用了变量替换 $y_2=1-d-x_2$. 继续计算上面的一串等式得

$$P\{X_{(i)}>X_{(i-1)}+d,i=2,3\}=3\int_0^{1-2d}(1-2d-x_1)^2\mathrm{d}x_1=3\int_0^{1-2d}y_1^2\mathrm{d}y_1=(1-2d)^3$$

因此，当 3 人在长度为 1 的区间上均匀地、独立地分布时，所求的没有 2 人彼此相距在 d 以内的概率等于$(1-2d)^3$，这里 $d\leqslant\frac{1}{2}$. 事实上，用同样的方法可证，当 n 人在单位区间上随机地分布时，当 $d\leqslant\frac{1}{n-1}$时，没有 2 人彼此相距在 d 以内的概率等于

$$[1-(n-1)d]^n$$

其证明留作练习. ■

第 j 个顺序统计量 $X_{(j)}$ 的密度函数可通过对联合密度函数——式(6-1)求积分或者用下面

的直接推理得到：要使 $X_{(j)}$ 等于 x，必须在 n 个值 $X_1, X_2, \cdots, X_n$ 中有 $j-1$ 个小于 x，有 $n-j$ 个大于 x，有一个等于 x. 任意给定 X_i 的一组值，其中 $j-1$ 个值小于 x，另一组 $n-j$ 个值大于 x，剩下的一个值等于 x 的概率密度为

$$[F(x)]^{j-1}[1-F(x)]^{n-j}f(x)$$

因此，由于把随机变量 $X_1, X_2, \cdots, X_n$ 分成上述三组共有

$$\binom{n}{j-1, n-j, 1} = \frac{n!}{(n-j)!\ (j-1)!}$$

种不同的分法，故 $X_{(j)}$ 的密度函数为

$$f_{X(j)}(x) = \frac{n!}{(n-j)!\ (j-1)!}[F(x)]^{j-1}[1-F(x)]^{n-j}f(x) \tag{6-2}$$

例 6b 设观测了大小为 $2n+1$ 的随机变量样本(即 $2n+1$ 个独立同分布的随机变量)，称其中第$(n+1)$个最小者为样本的中位数. 若从$(0,1)$上均匀分布的随机变量中，观测到大小为 3 的样本，试求此样本的中位数在$\frac{1}{4}$和$\frac{3}{4}$之间的概率.

解 由式(6-2)知，$X_{(2)}$ 的密度函数为

$$f_{X_{(2)}}(x) = \frac{3!}{1!\ 1!}x(1-x) \quad 0<x<1$$

275

因此

$$P\left\{\frac{1}{4} < X_{(2)} < \frac{3}{4}\right\} = 6\int_{1/4}^{3/4} x(1-x)\,\mathrm{d}x = 6\left\{\frac{x^2}{2} - \frac{x^3}{3}\right\}\Bigg|_{x=1/4}^{x=3/4} = \frac{11}{16}$$ ■

对式(6-2)求积分可得 $X_{(j)}$ 的累积分布函数，即

$$F_{X_{(j)}}(y) = \frac{n!}{(n-j)!(j-1)!}\int_{-\infty}^{y}[F(x)]^{j-1}[1-F(x)]^{n-j}f(x)\,\mathrm{d}x \tag{6-3}$$

然而，$F_{X(j)}(y)$也可以根据如下事实直接推出：第 j 个顺序统计量小于或等于 y 当且仅当 $X_1, X_2, \cdots, X_n$ 中至少有 j 个小于或等于 y. 由于 X_i 中小于或等于 y 的个数是以 $n, p=F(y)$为参数的二项随机变量，因而

$$\begin{aligned} F_{X_{(j)}}(y) &= P\{X_{(j)} \leqslant y\} = P\{X_i \text{ 中至少有 } j \text{ 个小于或等于 } y\} \\ &= \sum_{k=j}^{n}\binom{n}{k}[F(y)]^k[1-F(y)]^{n-k} \end{aligned} \tag{6-4}$$

如果在式(6-3)及式(6-4)中，我们取 F 为$(0, 1)$上的均匀分布函数[即 $f(x)=1, 0<x<1$]，则可得到一个有趣的解析恒等式

$$\sum_{k=j}^{n}\binom{n}{k}y^k(1-y)^{n-k} = \frac{n!}{(n-j)!(j-1)!}\int_0^y x^{j-1}(1-x)^{n-j}\,\mathrm{d}x \qquad 0 \leqslant y \leqslant 1 \tag{6-5}$$

采用与推导式(6-2)相类似的方法可以证明，当 $i<j$ 时，对一切 $x_i<x_j$，顺序统计量 $X_{(i)}$ 和 $X_{(j)}$ 的联合密度函数为

$$\begin{aligned} &f_{X_{(i)}, X_{(j)}}(x_i, x_j) \\ &= \frac{n!}{(i-1)!\ (j-i-1)!\ (n-j)!}[F(x_i)]^{i-1}[F(x_j)-F(x_i)]^{j-i-1}[1-F(x_j)]^{n-j}f(x_i)f(x_j) \end{aligned} \tag{6-6}$$

例 6c（随机样本的极差分布） 假定观测 n 个独立同分布的随机变量 $X_1, X_2, \cdots, X_n$，由 $R = X_{(n)} - X_{(1)}$ 定义的随机变量 R 称为 $X_1, X_2, \cdots, X_n$ 的极差(range)，如果随机变量 X_i 有分布函

276 数 F 及密度函数 f，则 R 的分布可由式(6-6)如下求出：对 $a\geqslant 0$，有

$$P\{R\leqslant a\}=P\{X_{(n)}-X_{(1)}\leqslant a\}=\iint_{x_n-x_1\leqslant a} f_{X_{(1)},X_{(n)}}(x_1,x_n)\mathrm{d}x_1\mathrm{d}x_n$$

$$=\int_{-\infty}^{\infty}\int_{x_1}^{x_1+a}\frac{n!}{(n-2)!}[F(x_n)-F(x_1)]^{n-2}f(x_1)f(x_n)\mathrm{d}x_n\mathrm{d}x_1$$

用变量替换 $y=F(x_n)-F(x_1)$，$\mathrm{d}y=f(x_n)\mathrm{d}x_n$ 得

$$\int_{x_1}^{x_1+a}[F(x_n)-F(x_1)]^{n-2}f(x_n)\mathrm{d}x_n=\int_0^{F(x_1+a)-F(x_1)}y^{n-2}\mathrm{d}y=\frac{1}{n-1}[F(x_1+a)-F(x_1)]^{n-1}$$

从而

$$P\{R\leqslant a\}=n\int_{-\infty}^{\infty}[F(x_1+a)-F(x_1)]^{n-1}f(x_1)\mathrm{d}x_1 \tag{6-7}$$

仅在少数特殊情况下，式(6-7)中的积分才能计算出来. 例如，当 $X_1,X_2,\cdots,X_n$ 都是在(0,1)上均匀分布时，对 $0<a<1$，由式(6-7)我们得

$$P\{R<a\}=n\int_0^1[F(x_1+a)-F(x_1)]^{n-1}f(x_1)\mathrm{d}x_1$$

$$=n\int_0^{1-a}a^{n-1}\mathrm{d}x_1+n\int_{1-a}^1(1-x_1)^{n-1}\mathrm{d}x_1$$

$$=n(1-a)a^{n-1}+a^n$$

对上式求导即得极差的密度函数为

$$f_R(a)=\begin{cases}n(n-1)a^{n-2}(1-a) & 0\leqslant a\leqslant 1\\ 0 & \text{其他}\end{cases}$$

也就是说，n 个独立的服从(0,1)上均匀分布的随机变量是参数为 $n-1,2$ 的 β 随机变量. ■

6.7 随机变量函数的联合概率分布

设 X_1,X_2 是联合连续型随机变量，其联合概率密度函数为 f_{X_1,X_2}. 随机变量 Y_1,Y_2 是 X_1,X_2 的函数，更明确地说，设 $Y_1=g_1(X_1,X_2)$，$Y_2=g_2(X_1,X_2)$，其中 g_1,g_2 是某些函数. 我们有 277 时需要计算 Y_1,Y_2 的联合分布函数.

假定函数 g_1,g_2 满足下列条件：

1. 方程组 $y_1=g_1(x_1,x_2)$，$y_2=g_2(x_1,x_2)$ 有唯一的以 y_1,y_2 表示 x_1,x_2 的解，即 $x_1=h_1(y_1,y_2)$，$x_2=h_2(y_1,y_2)$.

2. 函数 g_1,g_2 在所有点 (x_1,x_2) 上有连续的偏导数，而且在所有点 (x_1,x_2) 上，这两个函数满足下面的 2×2 阶行列式：

$$J(x_1,x_2)=\begin{vmatrix}\dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2}\\[2mm] \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2}\end{vmatrix}\equiv\frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2}-\frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1}\neq 0$$

在上面两个条件下，可证随机变量 Y_1,Y_2 是联合连续型的，其联合密度函数为

$$f_{Y_1,Y_2}(y_1,y_2)=f_{X_1,X_2}(x_1,x_2)|J(x_1,x_2)|^{-1} \tag{7-1}$$

其中 $x_1=h_1(y_1,y_2)$，$x_2=h_2(y_1,y_2)$.

式(7-1)的证明可按如下思路进行：

$$P\{Y_1\leqslant y_1,Y_2\leqslant y_2\}=\iint_{\substack{(x_1,x_2):\\ g_1(x_1,x_2)\leqslant y_1\\ g_2(x_1,x_2)\leqslant y_2}} f_{X_1,X_2}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2 \tag{7-2}$$

将式(7-2)关于 y_1, y_2 求导，则可得 Y_1, Y_2 的联合密度函数．求导的结果与式(7-1)的右边是相等的，这是高等微积分的一个习题，本书不加以证明．

例 7a 设 X_1, X_2 为具有概率密度函数 f_{X_1,X_2} 的联合连续型随机变量．令 $Y_1=X_1+X_2$，$Y_2=X_1-X_2$．试用 f_{X_1,X_2} 表示出 Y_1, Y_2 的联合密度函数．

解 令 $g_1(x_1,x_2)=x_1+x_2, g_2(x_1,x_2)=x_1-x_2$，则

$$J(x_1,x_2)=\begin{vmatrix}1 & 1\\ 1 & -1\end{vmatrix}=-2$$

又因方程组 $y_1=x_1+x_2, y_2=x_1-x_2$ 有解 $x_1=\frac{y_1+y_2}{2}, x_2=\frac{y_1-y_2}{2}$，故由式(7-1)可得所求的密度函数为

$$f_{Y_1,Y_2}(y_1,y_2)=\frac{1}{2}f_{X_1,X_2}\left(\frac{y_1+y_2}{2},\frac{y_1-y_2}{2}\right)$$

278

例如，若 X_1, X_2 是独立的且在(0,1)上有均匀分布的随机变量，则

$$f_{Y_1,Y_2}(y_1,y_2)=\begin{cases}\frac{1}{2} & 0\leqslant y_1+y_2\leqslant 2, 0\leqslant y_1-y_2\leqslant 2\\ 0 & \text{其他}\end{cases}$$

若 X_1, X_2 是独立的且分别以 λ_1, λ_2 为参数的指数随机变量，则

$$f_{Y_1,Y_2}(y_1,y_2)=\begin{cases}\frac{\lambda_1\lambda_2}{2}\exp\left\{-\lambda_1\left(\frac{y_1+y_2}{2}\right)-\lambda_2\left(\frac{y_1-y_2}{2}\right)\right\} & y_1+y_2\geqslant 0, y_1-y_2\geqslant 0\\ 0 & \text{其他}\end{cases}$$

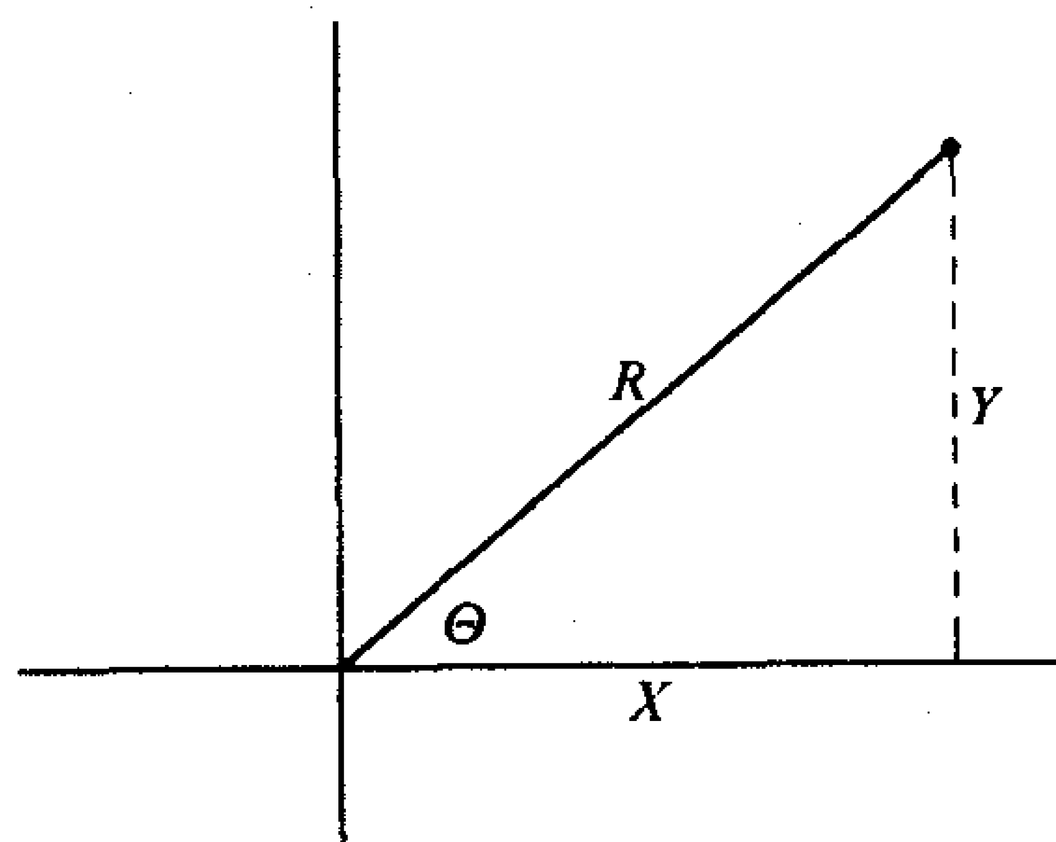
最后，如果 X_1, X_2 是独立的标准正态随机变量，则

$$f_{Y_1,Y_2}(y_1,y_2)=\frac{1}{4\pi}e^{-[(y_1+y_2)^2/8+(y_1-y_2)^2/8]}=\frac{1}{4\pi}e^{-(y_1^2+y_2^2)/4}$$

$$=\frac{1}{\sqrt{4\pi}}e^{-y_1^2/4}\frac{1}{\sqrt{4\pi}}e^{-y_2^2/4}$$

因此，我们不仅得到(与命题 3.2 一致)X_1+X_2，X_1-X_2 是均值为 0、方差为 2 的正态随机变量，而且还得到一个有趣的结果，即这两个随机变量是独立的．(事实上，可以证明，如果 X_1，X_2 是独立的随机变量且有共同的分布函数 F，当 F 是正态分布函数时，X_1+X_2 与 X_1-X_2 独立．) ■

例 7b 令(X,Y)代表平面内的一个随机点，假定直角坐标 X,Y 是独立的标准正态随机变量．我们感兴趣的是 R,Θ 的联合分布，其中 R,Θ 分别表示该点的极坐标(见图 6-4)．



•=随机点

$(X,Y)=R,\Theta$

图 6-4

令 $r=g_1(x,y)=\sqrt{x^2+y^2}, \theta=g_2(x,y)=\tan^{-1}y/x$，可以得到

$$\frac{\partial g_1}{\partial x}=\frac{x}{\sqrt{x^2+y^2}} \qquad \frac{\partial g_1}{\partial y}=\frac{y}{\sqrt{x^2+y^2}}$$

$$\frac{\partial g_2}{\partial x}=\frac{1}{1+(y/x)^2}\left(\frac{-y}{x^2}\right)=\frac{-y}{x^2+y^2}$$

$$\frac{\partial g_2}{\partial y}=\frac{1}{x[1+(y/x)^2]}=\frac{x}{x^2+y^2}$$

因此

$$J(x,y)=\frac{x^2}{(x^2+y^2)^{3/2}}+\frac{y^2}{(x^2+y^2)^{3/2}}=\frac{1}{\sqrt{x^2+y^2}}=\frac{1}{r}$$

X,Y 的联合密度函数为

$$f(x,y)=\frac{1}{2\pi}\mathrm{e}^{-(x^2+y^2)/2}$$

我们可以看到 $R=\sqrt{x^2+y^2}$，$\Theta=\tan^{-1}y/x$ 的联合密度函数为

$$f(r,\theta)=\frac{1}{2\pi}r\mathrm{e}^{-r^2/2}\qquad 0<\theta<2\pi,0<r<\infty$$

因为这个联合密度函数正好为 R,Θ 的边缘密度函数的乘积，因此 R,Θ 是独立的随机变量，且 Θ 服从$(0,2\pi)$上的均匀分布，R 服从瑞利(Rayleigh)分布，密度函数为

$$f(r)=r\mathrm{e}^{-\frac{r^2}{2}}\qquad 0<r<\infty$$

(因此，举例来说，当一个人向平面内的目标射击时，如果水平偏差和垂直偏差是独立的标准正态分布，则上述偏差的绝对值服从瑞利分布.)

上面的结果是十分有趣的，因为我们事先并不能完全肯定当一个随机向量的坐标是独立的标准正态随机变量时，其方向角不仅服从均匀分布，而且还与向量距原点的距离是独立的.

如果我们希望求得 R^2,Θ 的联合分布，则作变换 $d=g_1(x,y)=x^2+y^2$，$\theta=g_2(x,y)=\tan^{-1}y/x$，其雅可比行列式为

$$J=\begin{vmatrix} 2x & 2y \\ \dfrac{-y}{x^2+y^2} & \dfrac{x}{x^2+y^2}\end{vmatrix}=2$$

可以看到

$$f(d,\theta)=\frac{1}{2}\mathrm{e}^{-d/2}\frac{1}{2\pi}\qquad 0<d<\infty,0<\theta<2\pi$$

279～280

因此，R^2,Θ 是独立的，其中 R^2 服从参数为 1/2 的指数分布. 但是当 $R^2=X^2+Y^2$ 时，根据定义，R^2 是自由度为 2 的卡方分布. 因此，我们得到参数为 1/2 的指数分布就是自由度为 2 的卡方分布.

上面的结果通过在均匀随机变量上进行适当的变换，可以用来模拟(或者生成)正态随机变量. 令 U_1,U_2 表示独立的随机变量且均服从$(0,1)$上的均匀分布. 首先通过考虑(X_1,X_2)的极坐标(R,Θ)的表达式，我们把 U_1,U_2 变换成两个独立的标准正态随机变量 X_1,X_2. 从上面可知，R^2,Θ 是独立的，此外 $R^2=X_1^2+X_2^2$ 服从参数为 $\lambda=\frac{1}{2}$ 的指数分布. 但是 $-2\log U_1$ 之所以有如此分布，是因为对 $x>0$，

$$P\{-2\log U_1<x\}=P\left\{\log U_1>-\frac{x}{2}\right\}=P\{U_1>\mathrm{e}^{-x/2}\}=1-\mathrm{e}^{-x/2}$$

同样，$2\pi U_2$ 是$(0,2\pi)$上的均匀随机变量，我们可以利用它去生成 Θ，也就是说，如果令

$$R^2=-2\log U_1$$
$$\Theta=2\pi U_2$$

则 R^2 可以看作是从原点开始的距离的平方，θ 是(X_1,X_2)的方向角. 而 $X_1=R\cos\Theta$，$X_2=R\sin\Theta$，由此得到

$$X_1=\sqrt{-2\log U_1}\cos(2\pi U_2)$$
$$X_2=\sqrt{-2\log U_1}\sin(2\pi U_2)$$

是独立的标准正态随机变量. ■

例 7c 设 X,Y 是独立的 Γ 随机变量，参数分别为(α,λ),(β,λ)，求 $U=X+Y$,$V=X/(X+Y)$的联合密度.

解 X,Y 的联合密度为

$$f_{X,Y}(x,y)=\frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}\frac{\lambda e^{-\lambda y}(\lambda y)^{\beta-1}}{\Gamma(\beta)}=\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)}e^{-\lambda(x+y)}x^{\alpha-1}y^{\beta-1}$$

281

现在，如果 $g_1(x,y)=x+y$,$g_2(x,y)=x/(x+y)$,则

$$\frac{\partial g_1}{\partial x}=\frac{\partial g_1}{\partial y}=1\qquad\frac{\partial g_2}{\partial x}=\frac{y}{(x+y)^2}\qquad\frac{\partial g_2}{\partial y}=-\frac{x}{(x+y)^2}$$

所以

$$J(x,y)=\begin{vmatrix}1 & 1\\ \dfrac{y}{(x+y)^2} & \dfrac{-x}{(x+y)^2}\end{vmatrix}=-\frac{1}{x+y}$$

最后，由方程 $u=x+y$,$v=x/(x+y)$有 $x=uv$,$y=u(1-v)$，可以得到

$$f_{U,V}(u,v)=f_{X,Y}[uv,u(1-v)]u=\frac{\lambda e^{-\lambda u}(\lambda u)^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)}\frac{v^{\alpha-1}(1-v)^{\beta-1}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

因此 $X+Y$,$X/(X+Y)$是独立的，其中 $X+Y$ 服从参数为$(\alpha+\beta,\lambda)$的 Γ 分布，$X/(X+Y)$服从参数为(α,β)的 β 分布. 上式也证明了 $B(\alpha,\beta)$的标准化密度因子为

$$B(\alpha,\beta)\equiv\int_0^1 v^{\alpha-1}(1-v)^{\beta-1}dv=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

上述结果是非常有趣的. 假设要进行 $n+m$ 项工作，各个工作之间相互独立，完成工作的时间服从参数为 λ 的指数分布. 假设由两个工人来完成这些工作. 工人Ⅰ来完成 $1,2,\cdots,n$ 项工作，工人Ⅱ来完成剩下的 m 项工作. 令 X,Y 分别代表工人Ⅰ、工人Ⅱ总的工作时间，则(根据上述结果，或者由例 3b)，X,Y 独立的 Γ 随机变量，参数分别为(n,λ),(m,λ). 由前面的结论得出：由工人Ⅰ所完成的工作的时间与完成全部 $n+m$ 项工作的时间(即 $X+Y$)之比服从参数为(n,m)的 β 分布. ■

当给定 n 个随机变量 $X_1,\cdots,X_n$ 的联合密度函数时，我们要计算 $Y_1,\cdots,Y_n$ 的联合密度函数，其中

282

$$Y_1=g_1(X_1,\cdots,X_n),Y_2=g_2(X_1,\cdots,X_n),\cdots,Y_n=g_n(X_1,\cdots,X_n)$$

方法是相同的. 也就是说，假定函数 g_i 有连续的偏导数，而且在所有的点$(x_1,\cdots,x_n)$雅可比行列式 $J(x_1,\cdots,x_n)\neq 0$,其中

$$J(x_1,\cdots,x_n)=\begin{vmatrix}\dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} & \cdots & \dfrac{\partial g_1}{\partial x_n}\\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2} & \cdots & \dfrac{\partial g_2}{\partial x_n}\\ \cdots & \cdots & \cdots & \cdots\\ \dfrac{\partial g_n}{\partial x_1} & \dfrac{\partial g_n}{\partial x_2} & \cdots & \dfrac{\partial g_n}{\partial x_n}\end{vmatrix}$$

进一步，我们假定方程 $y_1=g_1(x_1,\cdots,x_n),y_2=g_2(x_1,\cdots,x_n),\cdots,y_n=g_n(x_1,\cdots,x_n)$ 有唯一解，比如说，$x_1=h_1(y_1,\cdots,y_n),\cdots,x_n=h_n(y_1,\cdots,y_n)$. 在这样的假定下，随机变量 Y_i 的联合密度函数为

$$f_{Y_1,\cdots,Y_n}(y_1,\cdots,y_n)=f_{X_1,\cdots,X_n}(x_1,\cdots,x_n)|J(x_1,\cdots,x_n)|^{-1} \tag{7-3}$$

其中 $x_i=h_i(y_1,\cdots,y_n),i=1,2,\cdots,n$.

例 7d 设 X_1,X_2,X_3 表示独立的标准正态随机变量. 若 $Y_1=X_1+X_2+X_3,Y_2=X_1-X_2,Y_3=X_1-X_3$，计算 Y_1,Y_2,Y_3 的联合密度函数.

解 令 $Y_1=X_1+X_2+X_3,Y_2=X_1-X_2,Y_3=X_1-X_3$，则这些变换的雅可比行列式为

$$J=\begin{vmatrix}1 & 1 & 1\\ 1 & -1 & 0\\ 1 & 0 & -1\end{vmatrix}=3$$

上述变换也等价于

$$X_1=\frac{Y_1+Y_2+Y_3}{3},X_2=\frac{Y_1-2Y_2+Y_3}{3},X_3=\frac{Y_1+Y_2-2Y_3}{3}$$

从式(7-3)可得

$$f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3)=\frac{1}{3}f_{X_1,X_2,X_3}\left(\frac{y_1+y_2+y_3}{3},\frac{y_1-2y_2+y_3}{3},\frac{y_1+y_2-2y_3}{3}\right)$$

因此，从

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3)=\frac{1}{(2\pi)^{3/2}}e^{-\sum_{i=1}^{3}x_i^2/2}$$

可以看到

283

$$f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3)=\frac{1}{3(2\pi)^{3/2}}e^{-Q(y_1,y_2,y_3)/2}$$

其中

$$\begin{aligned}Q(y_1,y_2,y_3)&=\left(\frac{y_1+y_2+y_3}{3}\right)^2+\left(\frac{y_1-2y_2+y_3}{3}\right)^2+\left(\frac{y_1+y_2-2y_3}{3}\right)^2\\&=\frac{y_1^2}{3}+\frac{2}{3}y_2^2+\frac{2}{3}y_3^2-\frac{2}{3}y_2y_3\end{aligned}$$

■

例 7c 设 $X_1,X_2,\cdots,X_n$ 是独立同分布的指数随机变量，参数为 λ. 令

$$Y_i=X_1+X_2+\cdots+X_i \qquad i=1,\cdots,n$$

(a) 求 $Y_1,\cdots,Y_n$ 的联合密度函数.

(b) 用(a)的结果求 Y_n 的密度.

解 (a) $Y_1=X_1,Y_2=X_1+X_2,\cdots,Y_n=X_1+\cdots+X_n$ 的雅可比行列式为

$$J=\begin{vmatrix}1 & 0 & 0 & 0 & \cdots & 0\\ 1 & 1 & 0 & 0 & \cdots & 0\\ 1 & 1 & 1 & 0 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & 1 & 1 & 1 & \cdots & 1\end{vmatrix}$$

因为只有行列式的第一列非零，我们有 $J=1$. 而 $X_1,\cdots,X_n$ 的联合密度函数为

$$f_{X_1,\cdots,X_n}(x_1,\cdots,x_n)=\prod_{i=1}^{n}\lambda e^{-\lambda x_i}\qquad 0<x_i<\infty, i=1,\cdots,n$$

因此，前面的变换相当于

$$X_1=Y_1, X_2=Y_2-Y_1,\cdots,X_i=Y_i-Y_{i-1},\cdots,X_n=Y_n-Y_{n-1}$$

从式(7-3)得 $Y_1,\cdots,Y_n$ 的联合密度函数为

$$\begin{aligned}f_{Y_1,\cdots,Y_n}(y_1,y_2\cdots,y_n)&=f_{X_1,\cdots,X_n}(y_1,y_2-y_1\cdots,y_i-y_{i-1},\cdots,y_n-y_{n-1})\\&=\lambda^n\exp\left\{-\lambda\left[y_1+\sum_{i=2}^{n}(y_i-y_{i-1})\right]\right\}\\&=\lambda^n e^{-\lambda y_n}\qquad 0<y_1,0<y_i-y_{i-1},i=2,\cdots,n\\&=\lambda^n e^{-\lambda y_n}\qquad 0<y_1<y_2<\cdots<y_n\end{aligned}$$

[284]

(b) 为了得到 Y_n 的边缘密度，我们一次对其他变量中的一个变量进行积分，有

$$f_{Y_2,\cdots,Y_n}(y_2,\cdots,y_n)=\int_0^{y_2}\lambda^n e^{-\lambda y_n}dy_1=\lambda^n y_2 e^{-\lambda y_n}\qquad 0<y_2<y_3<\cdots<y_n$$

继续下去给出

$$f_{Y_3,\cdots,Y_n}(y_3,\cdots,y_n)=\int_0^{y_3}\lambda^n y_2 e^{-\lambda y_n}dy_2=\lambda^n\frac{y_3^2}{2}e^{-\lambda y_n}\qquad 0<y_3<y_4<\cdots<y_n$$

再次积分给出

$$f_{Y_4,\cdots,Y_n}(y_4,\cdots,y_n)=\lambda^n\frac{y_4^3}{3!}e^{-\lambda y_n}\qquad 0<y_4<y_5<\cdots<y_n$$

按这种方式继续下去最后得到

$$f_{Y_n}(y_n)=\lambda^n\frac{y_n^{n-1}}{(n-1)!}e^{-\lambda y_n}\qquad 0<y_n$$

这与例 3b 得到的结果是一致的，证明了 $X_1+\cdots+X_n$ 是参数为 n,λ 的 Γ 随机变量. ■

*6.8 可交换随机变量

随机变量 $X_1,X_2,\cdots,X_n$ 是可交换的(exchangeable)，如果对整数 $1,\cdots,n$ 的任意排列 $i_1,\cdots,i_n$ 以及所有的 $x_1,\cdots,x_n$，有

$$P\{X_{i_1}\leqslant x_1,X_{i_2}\leqslant x_2,\cdots,X_{i_n}\leqslant x_n\}=P\{X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n\}$$

也就是说，如果它们的联合分布函数不论以何种顺序排列都是相同的，则称 n 个随机变量是可交换的.

离散型随机变量是可交换的，如果对所有的值 $x_1,\cdots,x_n$ 以及所有的排列 $i_1,\cdots,i_n$，有

$$P\{X_{i_1}=x_1,X_{i_2}=x_2,\cdots,X_{i_n}=x_n\}=P\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$$

这等价于 $p\{x_1,x_2,\cdots,x_n\}=P\{X_1=x,X_2=x_2,\cdots,X_n=x_n\}$ 是向量$(x_1,\cdots,x_n)$的对称函数，也就意味着当向量值的排列给定时，概率值不会改变.

例 8a 假定某罐中有 n 个球，其中有 k 个球是有记号的，每次从中无放回地取出 1 个球，且每次抽取是等可能的. 令

$$X_i=\begin{cases}1 & \text{抽到第 } i(i=1,2,\cdots,n)\text{个有记号的球}\\0 & \text{否则}\end{cases}$$

[285]

我们将证明随机变量 $X_1,\cdots,X_n$ 是可交换的. 令$(x_1,\cdots,x_n)$是包含 k 个 1 和 $n-k$ 个 0 的向量.

然而，在考虑点$(x_1,\cdots,x_n)$的联合质量函数的计算之前，我们首先通过考虑一给定的向量来得到一些结果，比如，考虑向量$(1,1,0,1,0,\cdots,0,1)$，其中含有k个1和$n-k$个0. 于是

$$p(1,1,0,1,0,\cdots,0,1)=\frac{k}{n}\frac{k-1}{n-1}\frac{n-k}{n-2}\frac{k-2}{n-3}\frac{n-k-1}{n-4}\cdots\frac{1}{2}\frac{1}{1}$$

这是因为第一次取到有记号球的概率为k/n，下一次取到有记号球的条件概率为$(k-1)/(n-1)$，再下一次取到无记号球的条件概率为$(n-k)/(n-2)$，等等. 根据同样的讨论可知，$p(x_1,\cdots,x_n)$可以表示成n个因子的乘积. 这些因子的分母项从n到1. 当向量$(x_1,\cdots,x_n)$的第i个位置为1时对应的分子项为$k-(i-1)$，而当第i个位置为0时对应的分子项为$n-k-(i-1)$. 于是，因为向量$(x_1,\cdots,x_n)$中包含k个1和$n-k$个0，所以我们得到

$$p(x_1,\cdots,x_n)=\frac{k!(n-k)!}{n!}\qquad x_i=0,1,\sum_{i=1}^{n}x_i=k$$

因为这是$(x_1,\cdots,x_n)$的对称函数，所以随机变量是可交换的. ■

注　得到上述公式的另一种方法是考虑n个球的联合概率质量函数，其中假定n个球是可区分的. 因为试验的结果依赖于这些球的排序问题，所以共有$n!$个等可能的结果. 其中有记号球与无记号球的排列的结果数等于有记号球的排列数与无记号球的排列数之积，即$k!(n-k)!$，由此可以得到上面的质量函数.

如果$X_1,X_2,\cdots,X_n$是可交换的，可知每个X_i有相同的概率分布. 比如，如果X,Y是可交换的离散型随机变量，则

$$P\{X=x\}=\sum_y P\{X=x,Y=y\}=\sum_y P\{X=y,Y=x\}=P\{Y=x\}$$

比如，从例8a可知无放回的情况下，取到的第i个球是有记号球的概率为k/n，因为凭直觉可以看到从n个球中取一个等同于选取第i个.

例8b　在例8a中，令Y_1表示取到第一个有记号球所抽取的次数，Y_2表示直到取到第二个有记号球时所增加的抽取次数，一般地，Y_i代表从取到第$(i-1)$个有记号球到取到第i个有记号球之间所间隔的次数，其中$i=1,\cdots,k$. 比如，如果$n=4,k=2,X_1=1,X_2=0,X_3=0,X_4=1$，则$Y_1=1,Y_2=3$. 现在，$Y_1=i_1,Y_2=i_2,\cdots,Y_k=i_k\Leftrightarrow X_{i_1}=X_{i_1+i_2}=\cdots X_{i_1+\cdots+i_k}=1$，其他$X_j=0$；于是，我们从$X_i$的联合质量函数得到

286

$$P\{Y_1=i_1,Y_2=i_2,\cdots,Y_k=i_k\}=\frac{k!\ (n-k)!}{n!}\qquad i_1+\cdots+i_k\leqslant n$$

因此，我们可以看到随机变量$Y_1,\cdots,Y_k$是可交换的. 例如，从一副洗好的牌中抽牌，直到第一张A出现所抽取的牌数的分布与第一张A出现后到下一张A又出现之间所抽取的牌数的分布相同. ■

例8c　该例是所谓波利亚(Polya)罐模型. 假定某罐中最初装有n个红球和m个蓝球. 每次随机地取一球，并记住其颜色，然后再加上一个同颜色的球一起放到罐中. 如果第i个球是红色的，则令$X_i=1$；如果第i个球是蓝色的，则令$X_i=0$，其中$i\geqslant 1$. 可用如下方法计算这些X_i的联合质量函数：

$$\begin{aligned}&P\{X_1=1,X_2=1,X_3=0,X_4=1,X_5=0\}\\&=\frac{n}{n+m}\frac{n+1}{n+m+1}\frac{m}{n+m+2}\frac{n+2}{n+m+3}\frac{m+1}{n+m+4}\\&=\frac{n(n+1)(n+2)m(m+1)}{(n+m)(n+m+1)(n+m+2)(n+m+3)(n+m+4)}\end{aligned}$$

和

$$
\begin{aligned}
& P\{X_1=0, X_2=1, X_3=0, X_4=1, X_5=1\} \\
& =\frac{m}{n+m} \frac{n}{n+m+1} \frac{m+1}{n+m+2} \frac{n+1}{n+m+3} \frac{n+2}{n+m+4} \\
& =\frac{n(n+1)(n+2)m(m+1)}{(n+m)(n+m+1)(n+m+2)(n+m+3)(n+m+4)}
\end{aligned}
$$

同理推出对包含 r 个 1 和 $k-r$ 个 0 的任意序列 $x_1, \cdots, x_k$，有

$$
P\{X_1=x_1, \cdots, X_k=x_k\}=\frac{n(n+1)\cdots(n+r-1)m(m+1)\cdots(m+k-r-1)}{(n+m)\cdots(n+m+k-1)}
$$

因此，可以看到对任意 k，随机变量 $X_1, \cdots, X_k$ 是可交换的. ■

287

最后一个例子是关于可交换的连续随机变量的问题.

例 8d 设 $X_1, X_2, \cdots, X_n$ 是独立的随机变量且服从(0,1)上的均匀分布，$X_{(1)}, \cdots, X_{(n)}$ 代表它们的顺序统计量. 也就是说，用 $X_{(j)}$ 表示 $X_1, X_2, \cdots, X_n$ 中第 j 小者. 另外，令

$$
\begin{aligned}
Y_1 &= X_{(1)} \\
Y_i &= X_{(i)}-X_{(i-1)} \qquad i=2, \cdots, n
\end{aligned}
$$

证明 $Y_1, \cdots, Y_n$ 是可交换的.

解 作变换

$$
y_1=x_1, \cdots, y_i=x_i-x_{i-1} \qquad i=2, \cdots n
$$

有

$$
x_i=y_1+\cdots+y_i \qquad i=1, \cdots, n
$$

很容易看出上面变换的雅可比行列式为 1，我们从式(7-3)得

$$
f_{Y_1, \cdots, Y_n}(y_1, y_2, \cdots, y_n)=f(y_1, y_1+y_2, \cdots, y_1+\cdots+y_n)
$$

其中 f 为顺序统计量的联合密度函数. 因此，从式(6-1)可得

$$
f_{Y_1, \cdots, Y_n}(y_1, y_2, \cdots, y_n)=n! \qquad 0<y_1<y_1+y_2<\cdots<y_1+\cdots+y_n<1
$$

或者等价地，

$$
f_{Y_1, \cdots, Y_n}(y_1, y_2, \cdots, y_n)=n! \qquad 0<y_i<1, i=1, \cdots, n, y_1+\cdots+y_n<1
$$

因为上述联合密度函数是 $y_1, \cdots, y_n$ 的对称函数，所以 $Y_1, \cdots, Y_n$ 是可交换的. ■

小结

随机变量对 X 和 Y 的联合累积概率分布函数定义为

$$
F(x, y)=P\{X \leqslant x, Y \leqslant y\} \qquad -\infty<x, y<\infty
$$

关于随机变量对的所有概率都可以通过 F 得到. 为了得到 X 和 Y 各自的概率分布函数，利用

$$
F_X(x)=\lim_{y \rightarrow \infty} F(x, y) \qquad F_Y(y)=\lim_{x \rightarrow \infty} F(x, y)
$$

如果 X 和 Y 都是离散型随机变量，那么它们的联合概率质量函数定义为

$$
p(i, j)=P\{X=i, Y=j\}
$$

288

各自的概率质量函数为

$$
P\{X=i\}=\sum_j p(i, j) \qquad P\{Y=j\}=\sum_i p(i, j)
$$

随机变量 X 和 Y 称为联合连续型随机变量，如果存在一个函数 $f(x, y)$，称为联合概率密

度函数，使得对任意的二维集合 C，有

$$P\{(X,Y)\in C\}=\iint_C f(x,y)\mathrm{d}x\mathrm{d}y$$

由前面的结果可知

$$P\{x<X<x+\mathrm{d}x,y<Y<y+\mathrm{d}y\}\approx f(x,y)\mathrm{d}x\mathrm{d}y$$

如果 X 和 Y 是联合连续型的，则它们都是连续型随机变量，且其密度函数为

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y \qquad f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}x$$

随机变量 X 和 Y 是独立的，若对一切集合 A,B，有

$$P\{X\in A,Y\in B\}=P\{X\in A\}P\{Y\in B\}$$

如果联合分布函数（或者离散情形下的联合概率质量函数，或者连续情形下的联合密度函数）可分解为两部分，一部分仅仅依赖于 x，另一部分仅仅依赖于 y，则 X 和 Y 是独立的.

一般情形下，随机变量 $X_1,\cdots,X_n$ 是独立的，如果对所有的实数集合 $A_1,\cdots,A_n$，有

$$P\{X_1\in A_1,\cdots,X_n\in A_n\}=P\{X_1\in A_1\}\cdots P\{X_n\in A_n\}$$

如果 X 和 Y 是独立的连续型随机变量，则它们的和的分布函数可以表示成以下形式：

$$F_{X+Y}(a)=\int_{-\infty}^{\infty}F_X(a-y)f_Y(y)\mathrm{d}y$$

如果 $X_i(i=1,\cdots,n)$ 是独立的正态随机变量，参数为 $\mu_i,\sigma_i^2,i=1,\cdots,n$，则 $\sum_{i=1}^{n}X_i$ 服从正态分布，其参数为 $\sum_{i=1}^{n}\mu_i,\sum_{i=1}^{n}\sigma_i^2$.

如果 $X_i(i=1,\cdots,n)$ 是独立的泊松随机变量，参数为 $\lambda_i,i=1,\cdots,n$，则 $\sum_{i=1}^{n}X_i$ 服从泊松分布，其参数为 $\sum_{i=1}^{n}\lambda_i$.

如果 X 和 Y 是离散型随机变量，则在给定 $Y=y$ 条件下 X 的条件概率质量函数为

289

$$P\{X=x|Y=y\}=\frac{p(x,y)}{p_Y(y)}$$

其中 p 是联合概率质量函数. 同样，如果 X 和 Y 是联合连续的，且联合概率密度函数为 f，则在给定 $Y=y$ 条件下 X 的条件概率密度函数为

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

一组独立同分布的随机变量的有序值 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 称为该组随机变量的顺序统计量. 如果随机变量是连续型的，密度函数为 f，则顺序统计量的联合密度函数为

$$f(x_1,\cdots,x_n)=n!\ f(x_1)\cdots f(x_n) \qquad x_1\leqslant x_2\leqslant\cdots\leqslant x_n$$

如果 $X_{i_1},\cdots,X_{i_n}$ 对 $1,\cdots,n$ 的每一个排列 $i_1,\cdots,i_n$ 有相同的联合分布，则称随机变量 $X_1,\cdots,X_n$ 是可交换的.

习题

1. 掷两个均匀的骰子，在下列情形下，求 X 和 Y 的联合概率质量函数.
   (a) X 是掷出的两个值中较大者，Y 是这两个值之和.

(b) X 是第一个骰子掷出的值，Y 是掷出的两个值中较大者.

(c) X 是掷出的两个值中较小者，Y 是掷出的两个值中较大者.

2. 假设一个罐里装有 5 个白球、8 个红球，从中无放回地抽取 3 个球. 如果所抽的第 i 个球是白色的，则令 $X_i=1$；如果所抽的第 i 个球是红色的，则令 $X_i=0$. 给出以下联合概率质量函数：

(a) X_1, X_2；　　(b) X_1, X_2, X_3.

3. 在习题 2 中，假定白球被编号，如果所抽的第 i 个球是白球，令 $Y_i=1$；如果所抽的第 i 个球是红球，令 $Y_i=0$. 求以下联合概率质量函数：

(a) Y_1, Y_2；　　(b) Y_1, Y_2, Y_3.

4. 假定在下一次抽球之前，将所抽的球再放回罐中，重复习题 2.

5. 假定在下一次抽球之前，将所抽的球再放回罐中，重复习题 3a.

6. 某包中有 5 个晶体管，其中 2 个是次品. 每次从中取出一个进行检验，直到所有的次品都检验出来为止. 记 N_1 为直到查出第一个次品时所需要的检验次数，N_2 为查出第一个次品后再查出第二个次品所需要的检验次数. 求 N_1 和 N_2 的联合概率质量函数.

7. 考虑独立的伯努利重复试验序列，其中每次试验成功的概率为 p. X_1 是第一次成功之前的失败次数，X_2 是在前两次成功之间的失败次数. 求 X_1 和 X_2 的联合概率质量函数.

8. 设 X 和 Y 的联合概率密度函数为

$$f(x,y)=c(y^2-x^2)e^{-y} \qquad -y\leqslant x\leqslant y, 0<y<\infty$$

(a) 求 c.

(b) 求 X 和 Y 的边缘密度函数.

(c) 求 $E[X]$.

9. 设 X 和 Y 的联合概率密度函数为

$$f(x,y)=\frac{6}{7}\left(x^2+\frac{xy}{2}\right) \qquad 0<x<1, 0<y<2$$

(a) 证明 $f(x,y)$ 确实是联合密度函数.

(b) 计算 X 的密度函数.

(c) 求 $P\{X>Y\}$.

(d) 求 $P\left\{Y>\frac{1}{2} \mid X<\frac{1}{2}\right\}$.

(e) 求 $E[X]$.

(f) 求 $E[Y]$.

10. 设 X 和 Y 的联合概率密度函数为

$$f(x,y)=e^{-(x+y)} \qquad 0\leqslant x<\infty, 0\leqslant y<\infty$$

求 (a) $P\{X<Y\}$；(b) $P\{X<a\}$.

11. 电视机商店的店主估计，来到商店的顾客中有 45%的人会买黑白电视机，15%的人会买彩色电视机，其余 40%的人什么都不买. 如果在某一天有 5 个顾客来到该店，问这一天他恰能售出 2 台黑白电视机和 1 台彩色电视机的概率是多少?

12. 在一个小时内，进入某药店的人数是参数为 $\lambda=10$ 的泊松分布随机变量. 已知在该时间段内已有 10 位女顾客进入店里，计算最多有 3 位男顾客进入药店的条件概率. 你做了怎样的假设?

13. 一个男人和一个女人约定下午 12:30 在某地见面. 如果男人到达的时间在 12:15 到 12:45 之间服从均匀分布. 女人独立到达，而且到达的时间在 12:00 到 13:00 之间服从均匀分布. 试求先到的人等待另一个人到达的时间不超过 5 分钟的概率. 如果男人先到概率又是多少?

14. 一辆救护车以恒定的速度沿着长度为 L 的道路往返. 在某时刻，道路上发生事故是均匀分布的. [即从某个固定端点开始行驶的距离服从$(0,L)$上的均匀分布.] 假定救护车在事故发生时的位置也是服

290

从均匀分布的，在独立的情况下，计算从事故现场到救护车的距离的分布.

15. 设随机向量(X,Y)在一平面区域 R 上是均匀分布的，如果对任一常数 c，联合密度为

$$f(x,y)=\begin{cases} c & \text{如果}(x,y)\in R \\ 0 & \text{其他} \end{cases}$$

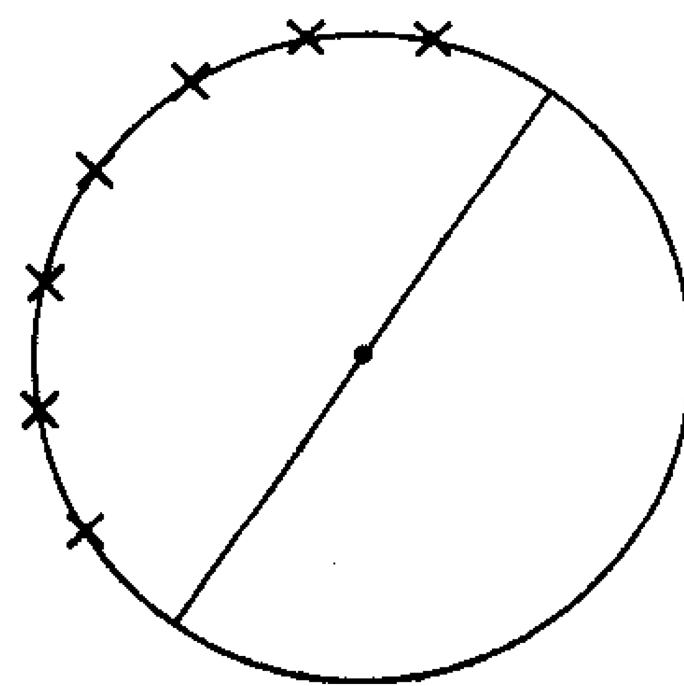
(a) 证明 $1/c$ 为区域 R 的面积.

假定(X,Y)在中心为$(0,0)$、边长为 2 的正方形区域上服从均匀分布.

(b) 证明 X 与 Y 是独立的且均服从$(-1, 1)$上的均匀分布.

(c) (X,Y)落在半径为 1、圆心为原点的圆的区域内的概率是多少？即求 $P\{X^2+Y^2\leqslant 1\}$.

16. 假定在一个圆的边界上独立地选取 n 个点，求全部点落在一个半圆周上的概率.（也就是求有一条直线过原点，使得所有的点都落在这条直线一边的概率.）



令 $P_1,\cdots,P_n$ 表示 n 个点，A 表示所有的点包含在某半圆周上的事件，A_i 表示所有的点落在从点 $P_i(i=1,\cdots,n)$开始沿着顺时针方向旋转 180°的半圆周上的事件.

(a) 用 A_i 表示 A. (b) A_i 是否互不相容？ (c) 求 $P(A)$.

17. 在一条直线 L 上随机地取三个点 X_1,X_2,X_3，试问 X_2 位于 X_1 和 X_3 之间的概率是多少？

18. 在长度为 L 的线段上随机地选两点，使这两点位于该线段的中点的两侧.（换句话说，两点 X 和 Y 是独立的随机变量，X 在$(0,L/2)$上均匀分布，而 Y 在$(L/2,L)$上均匀分布.）试求这两点间的距离大于 $L/3$ 的概率.

19. 在 18 题中，求从 0 到 X、从 X 到 Y 及从 Y 到 L 的三条线段能构成三角形的三条边的概率.（注意，当三条线段中任一线段的长度小于其余两线段的长度之和时，三条线段能构成三角形.）

20. 设 X 和 Y 的联合密度函数为

$$f(x,y)=\begin{cases} xe^{-(x+y)} & x>0,y>0 \\ 0 & \text{其他} \end{cases}$$

问 X 与 Y 是否独立？如果

$$f(x,y)=\begin{cases} 2 & 0<x<y,0<y<1 \\ 0 & \text{其他} \end{cases}$$

情况又如何？

21. 令

$$f(x,y)=\begin{cases} 24xy & 0\leqslant x\leqslant 1,0\leqslant y\leqslant 1,0\leqslant x+y\leqslant 1 \\ 0 & \text{其他} \end{cases}$$

(a) 证明 $f(x,y)$是联合概率密度函数. (b) 求 $E[X]$. (c) 求 $E[Y]$.

22. 设 X 和 Y 的联合密度函数为

$$f(x,y)=\begin{cases} x+y & 0<x<1,0<y<1 \\ 0 & \text{其他} \end{cases}$$

(a) 问 X 与 Y 是否独立？ (b) 求 X 的密度函数. (c) 求 $P\{X+Y<1\}$.

23. 设 X 和 Y 的联合密度函数为

$$f(x,y)=\begin{cases} 12xy(1-x) & 0<x<1,0<y<1 \\ 0 & \text{其他} \end{cases}$$

(a) 问 X 与 Y 是否独立？ (b) 求 $E[X]$. (c) 求 $E[Y]$.

(d) 求 $\mathrm{Var}(X)$. (e) 求 $\mathrm{Var}(Y)$.

24. 考虑独立的试验，试验结果为 $i(i=0,1,\cdots k)$的概率为 p_i，$\sum_{i=0}^{k} p_i = 1$. 令 N 表示试验的次数，不能为 0，X 表示试验结果.

291 ~ 292

(a) 求 $P\{N=n\},n\geqslant 1$. (b) 求 $P\{X=j\},j=1,\cdots,k$.

(c) 证明 $P\{N=n,X=j\}=P\{N=n\}P\{X=j\}$.

(d) 从直觉上看 N 与 X 独立吗? (e) 从直觉上看 X 与 N 独立吗?

25. 设有 10^6 个人要到某服务站，他们来到的时间(单位：小时)是独立随机变量，而且每个人来到的时间都服从$(0,10^6)$上的均匀分布. 令 N 表示他们当中在第一个小时内来到服务站的人数，试求 $P\{N=i\}$ 的近似值.

26. 设 A,B,C 为独立随机变量，在$(0,1)$上都服从均匀分布. 试问：

(a) A,B,C 的联合累积分布函数是什么?

(b) 方程 $Ax^2+Bx+C=0$ 的根全是实根的概率是多少?

27. 设 X 在$(0,1)$上服从均匀分布，Y 服从以 $\lambda=1$ 为参数的指数分布，又设 X 与 Y 独立，求以下概率分布：

(a) $Z=X+Y$； (b) $Z=X/Y$.

28. 设 X_1,X_2 是分别以 λ_1,λ_2 为参数的独立指数随机变量，试求 $Z=X_1/X_2$ 的分布，并计算 $P\{X_1<X_2\}$.

29. 当电流 I(单位：安培)通过电阻 R(单位：欧姆)时，所产生的功率为 $W=I^2R$(单位：瓦). 如果 I 和 R 为独立随机变量，其密度函数分别是

$$f_I(x)=6x(1-x) \qquad 0\leqslant x\leqslant 1$$
$$f_R(x)=2x \qquad 0\leqslant x\leqslant 1$$

试求 W 的密度函数.

30. 某杂志在一页中出现排版错误的数学期望为 0.2.

(a) 求一篇 10 页的文章中出现 0 个错误的概率.

(b) 求一篇 10 页的文章中出现 2 个或者更多错误的概率.

解释原因.

31. 每月商业航线出现航空事故的平均数为 2.2，求下列概率：

(a) 下个月出现 2 次以上事故.

(b) 下两个月出现 4 次以上事故.

(c) 下三个月出现 5 次以上事故.

解释原因.

32. 某饭店一个星期的总收入服从均值为 2200 美元、方差为 230 美元的正态分布. 求下列概率：

(a) 下两个星期总收入超过 5000 美元.

(b) 下三个星期里至少两星期的周收入超过 2000 美元.

你做了怎样的独立性假设?

33. 吉尔的保龄球分数近似服从均值为 170、标准差为 20 的正态分布，杰克的保龄球分数近似服从均值为 160、标准差为 15 的正态分布. 如果吉尔与杰克各打一局，并假定他们的分数是独立随机变量，求下列概率的近似值：

(a) 杰克的分数高.

(b) 他们的总分超过 350 分.

34. 美国国家健康统计中心的研究表明，25.2%的男性和 23.6%的女性从来不吃早饭. 假定选择 200 个男性和 200 个女性的随机样本. 求下列概率的近似值：

(a) 这 400 人中至少有 110 人从来不吃早饭.

(b) 女性从来不吃早饭的人数大于男性从来不吃早饭的人数.

35. 在习题 2 中，计算下列条件下 X_1 的条件概率质量函数：

293

(a) $X_2=1$;　(b) $X_2=0$.

36. 在习题 4 中，计算下列条件下 X_1 的条件概率质量函数：

(a) $X_2=1$;　(b) $X_2=0$.

37. 在习题 3 中，计算下列条件下 Y_1 的条件概率质量函数：

(a) $Y_2=1$;　(b) $Y_2=0$.

38. 在习题 5 中，计算下列条件下 Y_1 的条件概率质量函数：

(a) $Y_2=1$;　(b) $Y_2=0$.

39. 从数集$\{1,2,3,4,5\}$中随机地选出一数 X. 再从不大于 X 的子集即$\{1,2,\cdots,X\}$中随机地选取一个数，记此数为 Y.

(a) 求 X 和 Y 的联合质量函数.

(b) 对 $i=1,2,3,4,5$ 求给定 $Y=i$ 的条件下 X 的条件质量函数.

(c) 问 X 与 Y 是否独立？为什么？

40. 掷两个骰子，令 X 和 Y 分别表示掷出的最大点数和最小点数. 计算给定 $X=i$ 条件下 Y 的条件质量函数，其中 $i=1,2,\cdots,6$. 问 X 与 Y 是否独立？为什么？

41. 设 X 和 Y 的联合概率质量函数为

$$p(1,1)=\frac{1}{8}\qquad p(1,2)=\frac{1}{4}\qquad p(2,1)=\frac{1}{8}\qquad p(2,2)=\frac{1}{2}$$

(a) 计算给定 $Y=i$ 条件下 X 的条件质量函数，其中 $i=1,2$.

(b) X 与 Y 是否独立？

(c) 计算 $P\{XY\leqslant 3\}$, $P\{X+Y>2\}$, $P\{X/Y>1\}$.

42. 设 X 和 Y 的联合密度函数为

$$f(x,y)=xe^{-x(y+1)}\qquad x>0,y>0$$

(a) 求给定 $Y=y$ 条件下 X 的条件密度函数以及给定 $X=x$ 条件下 Y 的条件密度函数.

(b) 求 $Z=XY$ 的密度函数.

43. X 和 Y 的联合密度函数为

$$f(x,y)=c(x^2-y^2)e^{-x}\qquad 0\leqslant x<\infty,-x\leqslant y\leqslant x$$

求给定 $X=x$ 条件下 Y 的条件分布.

44. 某保险公司假定每个人都有一个事故参数，并且事故参数为 λ 的某人在一年内发生事故的次数服从参数为 λ 的泊松分布. 同时假定新参加保险的人的参数值是参数为 s,α 的 Γ 随机变量. 如果一个新参加保险的人在头一年发生 n 次事故，求该人的事故参数的条件密度函数. 同时，确定他下一年发生事故的期望值.

45. 设 X_1,X_2,X_3 是独立的且在(a,b)上有均匀分布的随机变量，试求其中的最大者大于其余两个之和的概率.

46. 一台机器的 5 个发动机中，只要有至少 3 个在运转，它就能有效地工作. 如果设这些发动机彼此独立地运转，而且每台运转的总时数是以 $f(x)=xe^{-x}(x>0)$为密度函数的随机变量，求此机器工作的持续时间的密度函数.

47. 设 3 辆卡车发生故障的地点随机地分布在长度为 L 的公路上，求其中没有 2 辆卡车彼此间的距离在 d 以内的概率，其中 $d\leqslant L/2$.

294

48. 考虑取自$(0,1)$上均匀分布的大小为 5 的样本. 计算其中位数落在区间$\left(\frac{1}{4},\frac{3}{4}\right)$内的概率.

49. 设 X_1,X_2,X_3,X_4,X_5 是以 λ 为参数的独立同分布指数随机变量.
    (a) 求 $P\{\min(X_1,\cdots,X_5)\leqslant a\}$. (b) 求 $P\{\max(X_1,\cdots,X_5)\leqslant a\}$.
50. 从密度函数为 $f(x)=2x(0<x<1)$ 的分布中，抽取大小为 2 的样本. 试求此样本极差的分布.
51. 设 X,Y 分别表示从一个半径为 1、圆心在坐标原点的圆内均匀地选取一点的横坐标与纵坐标，即它们的联合密度函数为

$$f(x,y)=\frac{1}{\pi} \qquad x^2+y^2\leqslant 1$$

    求极坐标 $R=(X^2+Y^2)^{1/2},\Theta=\tan^{-1}Y/X$ 的联合密度函数.
52. 设 X,Y 独立的随机变量且均服从(0,1)上的均匀分布，求 $R=\sqrt{X^2+Y^2},\Theta=\tan^{-1}Y/X$ 的联合密度函数.
53. 设 U 是$(0,2\pi)$上均匀分布的随机变量，Z 与 U 独立，且为速率为 1 的指数随机变量，定义 X,Y 如下：

$$X=\sqrt{2Z}\cos U$$
$$Y=\sqrt{2Z}\sin U$$

    直接证明(不用例 7b 的结果)X,Y 是独立的标准正态随机变量.
54. 设 X 和 Y 的联合密度函数为

$$f(x,y)=\frac{1}{x^2y^2} \qquad x\geqslant 1,y\geqslant 1$$

    (a) 计算 $U=XY$, $V=X/Y$ 的联合密度函数.
    (b) U 及 V 的边缘密度函数是什么?
55. 设 X 和 Y 均为(0,1)上均匀分布的独立同分布随机变量. 试求下列联合密度函数：
    (a) $U=X+Y$, $V=X/Y$.
    (b) $U=X$, $V=X/Y$.
    (c) $U=X+Y$, $V=X/(X+Y)$.
56. 设 X 和 Y 是独立的指数随机变量且参数均为 $\lambda=1$，重新解第 55 题.
57. 设 X_1,X_2 是独立的指数随机变量且参数均为 λ，求 $Y_1=X_1+X_2,Y_2=e^{X_1}$ 的联合密度函数.
58. 设 X,Y,Z 是独立的随机变量，且有相同的密度函数 $f(x)=e^{-x},0<x<\infty$. 试求 $U=X+Y$, $V=X+Z$, $W=Y+Z$ 的联合分布.
59. 在例 8b 中，令 $Y_{k+1}=n+1-\sum_{i=1}^{k}Y_i$，证明 $Y_1,\cdots,Y_k,Y_{k+1}$是可交换的. 如果在抽取中考虑到抽取的次序，则注意 Y_{k+1}为必须观察到一个有记号的球所抽取球的个数.
60. 一个罐里有 n 个球，分别编号为 $1,\cdots,n$，在其中随机地抽取 k 个. 如果抽到的是 i 号球，则 X_i 等于 1，否则 X_i 等于 0. 证明 $X_1,\cdots,X_n$ 是可交换的.

295

理论练习

1. 证明式(1-2).
2. 假设发生在某指定时间内的事件的个数是以 λ 为参数的泊松随机变量. 如果每一个事件相互独立地被归类，而被归入第 i 类的概率为 $p_i,i=1,\cdots,n,\sum p_i=1$. 试证发生第 i 类事件的个数是以 λp_i 为参数的独立泊松随机变量.
3. 给出一个用蒲丰投针问题估计 π 值的算法. 令人惊奇的是，这曾一度是计算 π 值的通用方法.
4. 当 $L>D$ 时，解蒲丰投针问题.

    答案：$\frac{2L}{\pi D}(1-\sin\theta)+2\theta/\pi$，其中 θ 满足 $\cos\theta=D/L$.

5. 设 X 和 Y 是独立的连续型正随机变量，试用 X 和 Y 的密度函数表示(a) $Z=X/Y$ 及(b) $Z=XY$ 的密度函数. 对 X 和 Y 都是指数随机变量的特殊情形，计算上述表达式.
6. 用解析方法(归纳法)证明，当 $X_i(i=1,2,\cdots,n)$ 是独立同分布的几何随机变量时，$X_1+\cdots+X_n$ 具有负二项分布. 另外，给出不需要任何计算的另一种证法.
7. (a) 设 X 服从 Γ 分布，参数为 (t,λ)，试问 $cX(c>0)$ 的分布是什么?

(b) 证明

$$\frac{1}{2\lambda}\chi_{2n}^2$$

是 Γ 分布，参数为 n,λ,n 是正整数，χ_{2n}^2 是自由度为 $2n$ 的卡方随机变量.
8. 设 X,Y 独立的连续型随机变量且失效率分别为 $\lambda_X(t),\lambda_Y(t)$，令 $W=\min(X,Y)$.

(a) 根据 X,Y 的分布求 W 的分布函数.

(b) 证明 W 的失效率函数 $\lambda_W(t)$ 为

$$\lambda_W(t)=\lambda_X(t)+\lambda_Y(t)$$

9. 设 $X_1,\cdots,X_n$ 是独立的指数随机变量，且具有共同的参数 λ，求 $\min(X_1,\cdots,X_n)$ 的分布.
10. 电池的寿命是独立的指数随机变量，参数为 λ. 一个手电筒需要两节电池才能工作. 如果一个人有一个手电筒及 n 节电池，问该手电筒工作时间的分布是什么?
11. 设 X_1,X_2,X_3,X_4,X_5 是独立的连续型随机变量，且具有共同的分布函数 F 和密度函数 f，记

$$I=P\{X_1<X_2<X_3<X_4<X_5\}$$

(a) 证明 I 不依赖于 F.

提示：把 I 写为 5 维积分，然后作变量替换 $u_i=F(x_i),i=1,\cdots,5$.

(b) 计算 I.

296

(c) 给出你关于(b) 的解答的直观解释.
12. 证明联合连续(离散)型随机变量 $X_1,\cdots,X_n$ 独立的充要条件是它们的联合概率密度(质量)函数 $f(x_1,\cdots,x_n)$ 可以写成

$$f(x_1,\cdots,x_n)=\prod_{i=1}^n g_i(x_i)$$

其中 $g_i(x)(i=1,\cdots,n)$ 是非负函数.
13. 在例 5c 中，我们计算了在前 $n+m$ 次试验中有 n 次成功的条件下成功概率的条件密度. 如果对指定 n 次成功的试验，条件密度会改变吗?
14. 设 X,Y 是独立的几何随机变量，有共同的参数 p.

(a) 不进行任何计算的前提下，你认为 $P\{X=i|X+Y=n\}$ 的值是什么?

提示：设想你连续抛一枚硬币，出现正面的概率为 p，如果在第 n 次抛硬币时出现第二个正面. 求第一个正面出现时抛的次数的概率质量函数.

(b) 验证你在(a) 中的推测.
15. 设 X,Y 是独立的二项随机变量，具有相同参数 n,p，用解析方法证明给定 $X+Y=m$ 的条件下 X 的条件分布是超几何分布. 另外，给出不需要进行任何计算的另外一种证法.

提示：假定抛 $2n$ 枚硬币，令 X 表示前 n 次抛出正面的次数，Y 表示后 n 次抛出正面的次数. 论证：已知抛出正面的总数等于 m 的条件下，前 n 次抛出正面的次数的分布与从 n 个白球和 n 个黑球中选出 m 个样本所含的白球数的分布相同.
16. 考虑只有 3 个可能结果的某个试验，结果 i 出现的概率为 $p_i,i=1,2,3$，设独立地重复该试验 n 次，令 $X_i(i=1,2,3)$ 表示结果 i 出现的次数. 求给定 $X_2=m$ 的条件下 X_1 的条件概率质量函数.
17. 设 X_1,X_2,X_3 是独立同分布的连续型随机变量，计算

(a) $P\{X_1>X_2|X_1>X_3\}$；　(b) $P\{X_1>X_2|X_1<X_3\}$；

(c) $P\{X_1>X_2|X_2>X_3\}$；　(d) $P\{X_1>X_2|X_2<X_3\}$.

18. 设 U 表示(0,1)上均匀分布的随机变量，在如下给定的条件下计算 U 的条件分布.

(a) $U>a$；　(b) $U<a$.

其中 $0<a<1$.

19. 设在某一天里，空气的湿度 W 是参数为 (t,β) 的 Γ 随机变量，即密度函数为 $f(w)=\beta e^{-\beta w}(\beta w)^{t-1}/\Gamma(t)$，$w>0$. 假定在给定 $W=w$ 的条件下，一天中事故发生的次数 N 服从均值为 w 的泊松分布. 证明在给定 $N=n$ 的条件下 W 服从参数为 $(t+n,\beta+1)$ 的 Γ 分布.

20. 设 W 是参数为 (t,β) 的 Γ 随机变量，假定在条件 $W=w$ 下，$X_1,\cdots,X_n$ 是独立的参数为 w 的指数随机变量. 证明在给定 $X_1=x_1,X_2=x_2,\cdots,X_n=x_n$ 的条件下，W 是参数为 $(t+n,\beta+\sum_{i=1}^{n}x_i)$ 的 Γ 随机变量.

297

21. 将 mn 个数排成一个 n 行 m 列的矩形阵列. 如果一个数是它所在行中的最小数且是它所在列中的最大数，则称该阵列包含一个鞍点(saddlepoint). 例如，在阵列

1	3	2
0	−2	6
0.5	12	3

中，第一行和第一列中的数字 1 就是一个鞍点. 鞍点的存在性在博弈论中是非常重要的. 考虑上述矩形阵列并且假设甲、乙二人独立进行如下游戏：甲从 $1,2,\cdots,n$ 中选择一个数，乙从 $1,2,\cdots,m$ 中选择一个数. 这些选择同时进行，如果甲选择了 i，乙选择了 j，那么甲从乙处赢的钱数由阵列的第 i 行第 j 列的数字所确定. 现在假定这个阵列包含一个鞍点，比如说是第 r 行第 k 列的那个数，记为 x_{rk}. 现在假如甲选择了第 r 行，那么他可以保证他赢的钱数至少为 x_{rk}(因为 x_{rk} 是第 r 行的最小值). 另一方面，假如乙选择了第 k 列，那么她保证她输的钱数不会超过 x_{rk}(因为 x_{rk} 是第 k 列的最大值). 于是，甲有了一种玩法，可以保证他赢的钱数至少为 x_{rk}，乙也有了一种玩法保证她输的钱数最多为 x_{rk}. 因此将这两个策略作为最优策略看来是合理的，并且对甲来说宣布游戏的值为 x_{rk}.

若上述矩形阵列中的 nm 个数字是从任意一个连续型分布中独立地选取，那么所选择的阵列中包含一个鞍点的概率是多少？

22. 称随机变量 X,Y 有二维正态分布，如果它们的联合密度函数为

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\times\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}$$

(a) 试证明给定 $Y=y$ 的条件下，X 的条件密度函数是以

$$\mu_x+\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\text{和}\ \sigma_x^2(1-\rho^2)$$

为参数的正态密度函数.

(b) 试证 X,Y 都是正态随机变量，其参数分别为 μ_x,σ_x^2 和 μ_y,σ_y^2.

(c) 证明当 $\rho=0$ 时，X,Y 独立.

23. 设 $F(x)$ 为累积分布函数. 证明当 n 为正整数时，(a) $F^n(x)$ 和 (b) $1-[1-F(x)]^n$ 也是累积分布函数. 提示：设 $X_1,\cdots,X_n$ 为独立的随机变量，有相同的分布函数 F，用 X_i 定义随机变量 Y,Z，使得 $P\{Y\leqslant x\}=F^n(x)$，$P\{Z\leqslant x\}=1-[1-F(x)]^n$.

298

24. 设 n 个人随机地分布在一条长 L 英里的路上，证明当 $D\leqslant L/(n-1)$ 时，没有两个人彼此相距小于 D 英里的概率为 $[1-(n-1)D/L]^n$. 如果 $D>L/(n-1)$，概率是多少？

25. 通过对式(6-4)求导来推出式(6-2).
26. 设一个大小为$(2n+1)$的样本来自$(0,1)$上的均匀分布，证明其中位数服从参数为$(n+1,n+1)$的β分布.
27. 验证给出了$X_{(i)},X_{(j)}$的联合密度的式(6-6).
28. 从密度函数为f的连续型随机变量中抽取大小为n的样本，计算此样本的极差的密度函数.
29. 设$X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$是$(0,1)$上均匀分布的$n$个独立的随机变量的顺序统计量，证明对$1\leqslant k\leqslant n+1$，有

$$P\{X_{(k)}-X_{(k-1)}>t\}=(1-t)^n$$

其中$X_{(0)}\equiv 0,X_{(n+1)}\equiv t$.
30. 设$X_1,\cdots,X_n$是一组独立同分布的连续型随机变量，分布函数为F，令$X_{(i)}(i=1,\cdots,n)$表示它们的有序值. 若X与$X_i(i=1,\cdots,n)$独立，且有相同的分布函数F，计算
(a) $P\{X>X_{(n)}\}$;　(b) $P\{X>X_{(1)}\}$;　(c) $P\{X_{(i)}<X<X_{(j)}\},1\leqslant i<j\leqslant n$.
31. 设$X_1,\cdots,X_n$是一组独立同分布的连续型随机变量，分布函数为F，密度函数为f. $M\equiv[X_{(1)}+X_{(n)}]/2$定义为最大值与最小值的均值，称为中程数. 证明其分布函数为

$$F_M(m)=n\int_{-\infty}^{m}[F(2m-x)-F(x)]^{n-1}f(x)\mathrm{d}x$$

32. 设$X_1,\cdots,X_n$是独立的且$(0,1)$上均匀分布的随机变量. 令$R=X_{(n)}-X_{(1)}$表示极差，$M=[X_{(1)}+X_{(n)}]/2$表示中程数. 计算R,M的联合密度函数.
33. 设X和Y是独立的标准正态随机变量. 计算

$$U=X \qquad V=\frac{X}{Y}$$

的联合密度函数. 然后运用你的结果证明X/Y服从柯西分布.

自测题与练习

1. 掷一个不均匀的骰子，出现奇数1,3,5的概率为C，出现偶数2,4,6的概率为$2C$，
(a) 求C.
(b) 掷这个骰子，若结果为偶数，记X为1，否则记X为0. 如果掷出的点数大于3，记Y为1，否则记Y为0. 求X和Y的联合质量函数.
现在假定独立地掷12次骰子.
(c) 求6点恰好出现两次的概率.
(d) 求其中4次出现1点或者2点、4次出现3点或4点、4次出现5点或6点的概率.
(e) 求至少有8次出现偶数的概率.

299

2. 随机变量X,Y,Z的联合概率质量函数为

$$p(1,2,3)=p(2,1,1)=p(2,2,1)=p(2,3,2)=\frac{1}{4}$$求 (a) $E[XYZ]$; (b) $E[XY+XZ+YZ]$.
3. 设X和Y的联合密度函数为

$$f(x,y)=C(y-x)\mathrm{e}^{-y} \qquad -y<x<y,0<y<\infty$$

(a) 求C.
(b) 求X的密度函数.
(c) 求Y的密度函数.
(d) 求$E[X]$.
(e) 求$E[Y]$.
4. 设X,Y,Z是等可能地取值为1或2的独立随机变量，求下列随机变量的概率质量函数.
(a) XYZ.

(b) $XY+XZ+YZ$.

(c) X^2+YZ.

5. 设 X 和 Y 是连续型随机变量，联合密度函数为

$$f(x,y)=\begin{cases}\dfrac{x}{5}+cy & 0<x<1,1<y<5\\ 0 & \text{否则}\end{cases}$$

其中 c 是常数.

(a) 求 c 的值.

(b) X 与 Y 独立吗?

(c) 求 $P\{X+Y>3\}$.

6. 设 X 和 Y 的联合密度函数为

$$f(x,y)=\begin{cases}xy & 0<x<1,0<y<2\\ 0 & \text{否则}\end{cases}$$

(a) X 与 Y 独立吗?

(b) 求 X 的密度函数.

(c) 求 Y 的密度函数.

(d) 求联合分布函数.

(e) 求 $E[Y]$.

(f) 求 $P\{X+Y<1\}$.

7. 考虑 2 个元件和 3 种冲击类型. 冲击类型 1 可以引起元件 1 失效，冲击类型 2 可以引起元件 2 失效，冲击类型 3 可以引起两个元件同时失效. 类型 1，2，3 的冲击时间是独立的指数随机变量，参数分别为 $\lambda_1,\lambda_2,\lambda_3$. 令 X_i 表示元件 i 失效的时间，其中 $i=1,2$. 随机变量 X_1,X_2 称为有二维联合指数分布. 求 $P\{X_1>s,X_2>t\}$.

8. 某分类广告的登记簿由 m 页组成，其中 m 是一个很大的数值. 每页的广告数量都是变化的，想知道每一页有多少个广告的唯一方法就是去数它们. 另外，假定页数非常多，没有办法数清楚总数，你的目的就是用一种方法保证广告登记簿中的每一个广告都等可能地被选择.

(a) 如果你在广告簿中随机地选择一页，然后随机地选择一个广告，能够实现你的目的吗? 为什么? 300

用 $n(i)$ 代表第 $i(i=1,\cdots,m)$ 页的广告数，假定数量是未知的，我们可以假定这些数都小于等于一个固定的值 n. 考虑下面选取广告的算法.

步骤 1：随机地选择一页，假设为第 X 页. 数第 X 页的广告数来确定 $n(X)$.

步骤 2：以概率 $n(X)/n$ 来“接受”第 X 页，如果接受了 X，则进行第三步，否则返回第一步.

步骤 3：在 X 页随机地选择一个广告.

称每通过第一步的算法为一次迭代. 比如，第一个随机选择的页数被拒绝，第二个被接受，则需要两次迭代去获得广告.

(b) 在一次迭代的情况下，接受第 i 页广告的概率有多大?

(c) 在一次迭代的情况下，接受一个广告的概率有多大?

(d) 在 k 次迭代的情况下，在最后的迭代中接受第 i 页上第 j 个广告的概率有多大?

(e) 依照上面的算法，得到第 i 页上第 j 个广告的概率有多大?

(f) 上述算法中，迭代次数的期望值是多少?

9. 自测题 8 中的算法的“随机”部分可以根据服从(0,1)上均匀分布的随机变量序列的生成值(称为随机数)来编写. 以$[x]$表示不大于 x 的最大整数，算法的步骤 1 可以编写如下：

步骤 1：产生一个服从(0,1)上均匀分布的随机变量 U. 令 $X=[mU]+1$，确定 $n(X)$ 的值.

(a) 解释以上为什么等价于自测题 8 中的步骤 1.

提示:X 的概率质量函数是什么?

(b) 以类似的方式编写算法的其余步骤.

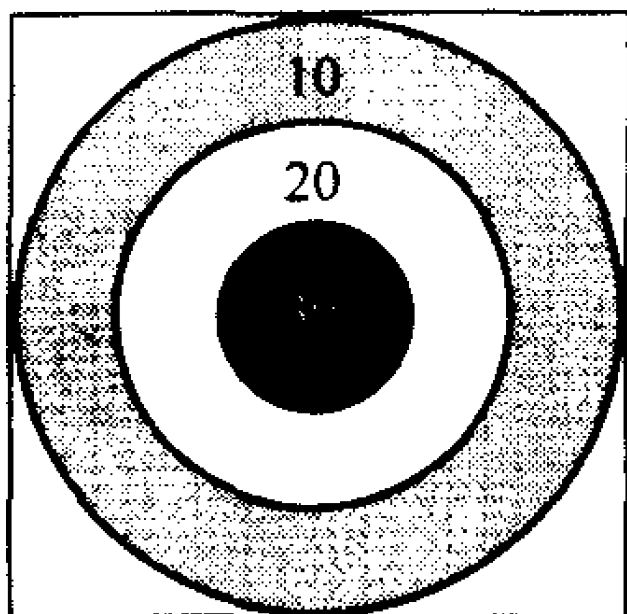
10. 设 $X_1, X_2, \cdots$ 是(0,1)上均匀分布的独立随机变量序列. 对固定的常数 c，定义随机变量 N 为

$$N=\min\{n: X_n>c\}$$

那么 N 与 X_N 独立吗? 也就是说,已知第一个随机变量的值大于 c 是否会影响这个随机变量出现的概率分布? 对你的结果给出直观的解释.

11. 某掷镖的靶是边长为 6 的正方形，里面有 3 个圆，半径分别为 1,2,3. 投在半径为 1 的投掷区域里得分为 30，投在半径为 1 的圆之外而在半径为 2 的圆之内得分为 20，投在半径为 2 的圆之外而在半径为 3 的圆之内得分为 10，投在半径为 3 的圆之外不得分. 假设每次投掷与前面的投掷是独立的，投掷的点在正方形内均匀分布，计算下列事件的概率：

301



(a) 一次投掷得 20 分.

(b) 一次投掷至少得 20 分.

(c) 一次投掷得 0 分.

(d) 一次投掷得分的数学期望.

(e) 前两次投掷都至少得 10 分.

(f) 两次投掷后总分为 30 分.

12. NBA 篮球赛提出的模型是假定两个队在比赛中有大体相同的积分，则在四分之一比赛中,主队的得分减去客队的得分近似为正态随机变量，均值为 1.5，方差为 6. 此外，模型假定四场四分之一比赛的得分差是独立的. 在这个模型的假定下,计算

(a) 主队胜出的概率是多少?

(b) 在某场比赛中主队半场落后 5 分的条件下，主队赢的条件概率是多少?

(c) 在第一场四分之一比赛结束时主队领先 5 分的条件下，主队赢的条件概率是多少?

13. 令 N 为参数为 p 的几何随机变量. 假定在给定 $N=n$ 的条件下 X 的条件分布是参数为 n,λ 的 Γ 分布. 求给定 $X=x$ 的条件下 N 的条件概率质量函数.

14. 设 X 和 Y 是(0,1)上独立的均匀分布的随机变量.

(a) 求 $U=X, V=X+Y$ 的联合密度函数.

(b) 使用(a) 得到的结果计算 V 的密度函数.

15. 你和其他 3 个人对一项工程投标，出价高的标将赢得该项工程. 如果你赢了，你计划立即以 10 千美元出售. 如果你认为别人的投标价是独立的，且在(7, 11)千美元上均匀分布，为得到最大期望利润你应该投多大的标?

16. 求 $X_1, X_2, \cdots, X_n$ 是 $1,2,\cdots,n$ 的一个排列的概率，其中 $X_1, X_2, \cdots, X_n$ 是独立的且

(a) 每一个 X_i 等可能地在 $1,2,\cdots,n$ 中任取一值.

(b) 每一个 X_i 有概率质量函数 $P\{X_i=j\}=p_j, j=1,\cdots,n$.

17. 设 $X_1, X_2, \cdots, X_n$ 与 $Y_1, Y_2, \cdots, Y_n$ 是独立的随机向量，且每个向量都是 k 个 1 和 $n-k$ 个 0 的随机排列. 也就是说，它们的联合概率质量函数为

$$P\{X_1=i_1,\cdots,X_n=i_n\}=P\{Y_1=i_1,\cdots,Y_n=i_n\}=\frac{1}{\binom{n}{k}}, i_j=0,1, \sum_{j=1}^{n} i_j=k$$

302

令

$$N=\sum_{i=1}^{n}|X_i-Y_i|$$

表示两个向量的坐标值不同的数目，M 表示满足 $X_i=1, Y_i=0$ 的 i 的数目.

(a) 写出 N 与 M 之间的关系式.

(b) M 的分布是什么?

(c) 求 $E[N]$.

(d) 求 $\text{Var}(N)$.

* 18. 设 $Z_1, Z_2, \cdots, Z_n$ 是独立的标准正态随机变量，令

$$S_j = \sum_{i=1}^{j} Z_i$$

(a) 给定 $S_k=y(k=1,\cdots,n)$ 条件下 S_n 的条件分布是什么?

(b) 证明对 $1\leqslant k\leqslant n$，给定 $S_n=x$ 的条件下 S_k 的条件分布是正态的，且均值为 xk/n，方差为 $k(n-k)/n$.

303

第7章　数学期望的性质

7.1　引言

本章我们将学习数学期望的其他性质．首先，回顾一下随机变量 X 的数学期望的定义：

$$E[X]=\sum_{x}xp(x)$$

此处 X 为离散型随机变量且有概率质量函数 $p(x)$．当 X 为连续型随机变量且有密度函数 $f(x)$时，有

$$E[X]=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x$$

因为 $E[X]$是 X 的所有可能取值的加权平均，所以如果 X 在 a 和 b 间取值，那么期望值也在此范围内取值．也就是说，如果

$$P\{a\leqslant X\leqslant b\}=1$$

那么

$$a\leqslant E[X]\leqslant b$$

为证明这个结果，假设 X 是一个离散型随机变量且 $P\{a\leqslant X\leqslant b\}=1$，这就意味着对于所有区间$[a,b]$以外的 x 均有 $p(x)=0$，故有

$$E[X]=\sum_{x:p(x)>0}xp(x)\geqslant\sum_{x:p(x)>0}ap(x)=a\sum_{x:p(x)>0}p(x)=a$$

同理可证 $E[X]\leqslant b$，故对离散型随机变量上式成立，同样对连续型随机变量也成立．

304

7.2　随机变量和的数学期望

对于二维情况，第 4 章命题 5.1 和第 5 章命题 2.1 给出了随机变量函数的数学期望的计算公式，假设 X 和 Y 为随机变量，g 为这两个变量的函数，我们有下面的结论．

命题 2.1　*如果 X 和 Y 有联合概率质量函数 $p(x,y)$，则*

$$E[g(X,Y)]=\sum_{y}\sum_{x}g(x,y)p(x,y)$$

如果 X 和 Y 有联合概率密度函数 $p(x,y)$，则

$$E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$$

例 2a　一次事故发生在一条长为 L 的道路上的点 X 服从均匀分布．在事故发生的时刻，救护车所在位置 Y 在这条道路上也服从均匀分布．假设 X 和 Y 相互独立，求事故发生点和救护车所在位置之间距离的期望值．

解　我们需要计算 $E[|X-Y|]$．因为 X 和 Y 的联合密度函数为

$$f(x,y)=\frac{1}{L^2}\quad 0<x<L,0<y<L$$

由命题 2.1 可得

$$E[|X-Y|]=\frac{1}{L^2}\int_0^L\int_0^L|x-y|\mathrm{d}y\mathrm{d}x$$

因为

$$\int_0^L |x-y|\,dy = \int_0^x (x-y)dy + \int_x^L (y-x)dy$$

305

$$= \frac{x^2}{2} + \frac{L^2}{2} - \frac{x^2}{2} - x(L-x) = \frac{L^2}{2} + x^2 - xL$$

因此

$$E[|X-Y|] = \frac{1}{L^2}\int_0^L \left(\frac{L^2}{2} + x^2 - xL\right)dx = \frac{L}{3}$$ ■

作为命题 2.1 的一个重要应用，假设 $E[X]$和 $E[Y]$均为有限值，且令 $g(X,Y)=X+Y$，那么，在连续情形下

$$E[X+Y] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f(x,y)dxdy$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dydx + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy$$
$$= \int_{-\infty}^{+\infty}xf_X(x)dx + \int_{-\infty}^{+\infty}yf_Y(y)dy = E[X]+E[Y]$$

在一般情形下上面结果同样成立，故如果 $E[X]$和 $E[Y]$均为有限值，则

$$E[X+Y]=E[X]+E[Y] \tag{2-1}$$

例 2b 假设对随机变量 X 和 Y，满足

$$X\geqslant Y$$

也就是说，对随机试验的任何结果，随机变量 X 的值大于等于随机变量 Y 的值，等价于 $X-Y\geqslant 0$，由此得出 $E[X-Y]\geqslant 0$，或等价地，

$$E[X]\geqslant E[Y]$$ ■

应用式(2-1)，简单的归纳可证：如果对所有的 $i=1,\cdots,n$，$E[X_i]$取有限值，那么

$$E[X_1+\cdots+X_n]=E[X_1]+\cdots+E[X_n] \tag{2-2}$$

以下几个例子可以说明式(2-2)是一个非常有用的公式.

例 2c(样本均值) 设 $X_1,\cdots,X_n$ 为有相同的分布函数 F 和相同的数学期望 μ 的独立同

306

分布的随机变量，我们称 $X_1,\cdots,X_n$ 为来自分布 F 的一个样本. 称

$$\overline{X} = \sum_{i=1}^{n}\frac{X_i}{n}$$

为样本均值(sample mean). 计算 $E[\overline{X}]$.

解

$$E[\overline{X}] = E\left[\sum_{i=1}^{n}\frac{X_i}{n}\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}X_i\right] = \frac{1}{n}\sum_{i=1}^{n}E[X_i]$$
$$= \mu \qquad \text{因为 } E[X_i]\equiv\mu$$

即样本均值的数学期望等于分布的数学期望 μ. 在统计学里样本均值常用来估计分布的数学期望. ■

例 2d(布尔不等式) 设 $A_1,\cdots,A_n$ 表示事件，并定义示性变量 $X_i(i=1,\cdots,n)$为

$$X_i=\begin{cases}1 & \text{如果 } A_i \text{ 发生}\\ 0 & \text{否则}\end{cases}$$

令 $X=\sum_{i=1}^{n}X_i$，因此 X 表示事件 A_i 发生的次数，最后令

$$Y=\begin{cases}1 & \text{若 } X\geqslant 1\\ 0 & \text{否则}\end{cases}$$

即如果至少有一个 A_i 发生则 Y 等于 1，否则为 0. 立即可得

$$X \geqslant Y$$

于是

$$E[X] \geqslant E[Y]$$

然而由于

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P(A_i)$$

307

并且

$$E[Y] = P\{\text{至少有一个 } A_i \text{ 发生}\} = P\left(\bigcup_{i=1}^{n} A_i\right)$$

因此得到布尔不等式，即

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} P(A_i)$$

■

下面三个例子说明式(2-2)可以用来计算二项随机变量、负二项随机变量、超几何随机变量的数学期望. 这些推导可以与第 4 章所用的方法进行比较.

例 2e(二项分布随机变量的数学期望) 设 X 是以 n 和 p 为参数的二项随机变量. 回忆一个随机变量表示在 n 次独立试验中成功的次数，每次成功的概率为 p，我们有

$$X = X_1 + X_2 + \cdots + X_n$$

其中

$$X_i = \begin{cases} 1 & \text{若第 } i \text{ 次试验成功} \\ 0 & \text{若第 } i \text{ 次试验不成功} \end{cases}$$

因此，X_i 是一个伯努利随机变量，且有数学期望 $E[X_i] = 1(p) + 0(1-p)$. 于是

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n] = np$$

■

例 2f(负二项随机变量的数学期望) 对于一系列独立试验，每次成功的概率为 p，试确定获得 r 次成功需要的试验次数的数学期望.

解 如果 X 表示获得 r 次成功所需要的试验次数，那么 X 为负二项随机变量，可以表示为

$$X = X_1 + \cdots + X_r$$

其中 X_1 表示获得第一次成功所需的试验次数，X_2 表示直到获得第二次成功另外需要进行的试验次数，X_3 表示直到获得第三次成功另外需要进行的试验次数，等等. 即，X_i 表示从第 $i-1$ 次成功到第 i 次成功所需的试验次数. 显然，每个随机变量 X_i 均是以 p 为参数的几何随机变量. 于是，由第 4 章例 9b，$E[X_i] = \frac{1}{p}$，$i = 1, 2, \cdots, r$；因此

$$E[X] = E[X_1] + \cdots + E[X_r] = \frac{r}{p}$$

■

308

例 2g(超几何随机变量的数学期望) 设某罐中装有 N 个球，其中白球有 m 个，从中取出 n 个球，求取出的白球数的数学期望.

解 设 X 表示取出的白球数，X 可表示为

$$X = X_1 + \cdots + X_m$$

其中

$$X_i = \begin{cases} 1 & \text{若第 } i \text{ 个白球被取出} \\ 0 & \text{若第 } i \text{ 个白球未被取出} \end{cases}$$

因为

$$E[X_i]=P\{X_i=1\}=P\{第\ i\ 个白球被取出\}=\frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}}=\frac{n}{N}$$

因此

$$E[X]=E[X_1]+\cdots+E[X_m]=\frac{mn}{N}$$

也可以用另外一种表示方法获得这个结果，即令

$$X=Y_1+\cdots+Y_n$$

其中

$$Y_i=\begin{cases}1 & 如果选择的第\ i\ 个球为白球\\0 & 如果选择的第\ i\ 个球不是白球\end{cases}$$

由于第 i 个球等可能地从 N 个球中取出，由此推出 $E[Y_i]=\frac{m}{N}$.

因此

$$E[X]=E[Y_1]+\cdots+E[Y_n]=\frac{nm}{N}$$ ■

例 2h(配对问题的数学期望) 有 N 个人将他们的帽子抛向屋子中央，将帽子充分混合后，每人随机地从中取出一顶，求刚好拿到自己帽子的人数的数学期望.

309

解 设 X 表示帽子和人刚好配对的人数，很容易计算 $E[X]$，令

$$X=X_1+X_2+\cdots+X_N$$

其中

$$X_i=\begin{cases}1 & 第\ i\ 个人选中自己的帽子\\0 & 第\ i\ 个人未选中自己的帽子\end{cases}$$

由于对每个 i，第 i 个人等可能地从 N 个帽子中取出一顶，

$$E[X_i]=P\{X_i=1\}=\frac{1}{N}$$

因此

$$E[X\}=E[X_1]+\cdots+E[X_N]=\left(\frac{1}{N}\right)N=1$$

所以平均来说,刚好有一个人取到自己的帽子. ■

例 2i 这个问题是由 18 世纪数学家伯努利提出并解决的. 假设一个坛子里放有 $2N$ 张卡片，其中有两个标有记号 1，两个标有记号 2，两个标有记号 3,等等，从中随机取出 m 张卡片，坛子里剩下的卡片中仍然成对的数目的期望值为多少?（有趣的是，伯努利曾用这个模型来研究 N 对夫妇中有 m 人去世后仍然成对的夫妻数的问题.）

解 对于 $i=1, 2, \cdots, N$，定义随机变量

$$X_i=\begin{cases}1 & 如果第\ i\ 对卡片还在坛子中\\0 & 否则\end{cases}$$

则

$$E[X_i]=P\{X_i=1\}=\frac{\binom{2N-2}{m}}{\binom{2N}{m}}=\frac{\frac{(2N-2)!}{m!\ (2N-2-m)!}}{\frac{(2N)!}{m!\ (2N-m)!}}=\frac{(2N-m)(2N-m-1)}{(2N)(2N-1)}$$

310

因此，所求的结果为

$$E[X_1+X_2+\cdots+X_N]=E[X_1]+\cdots+E[X_N]=\frac{(2N-m)(2N-m-1)}{2(2N-1)}$$ ■

例 2j（票券收集问题） 假设有 N 种不同类型的票券，每次获得其中的一种票券，无论它属于哪种类型都是等可能的.

(a) 现获得 n 张票券，试求其中所含的不同类型数的期望值.

(b) 为获得每种类型至少有一张的整套票券，试求所需收集的票券数的数学期望.

解 (a) 令 X 表示这 n 张票券中所含的不同类型数，我们用表达式

$$X=X_1+\cdots+X_N$$

来计算 $E[X]$，其中

$$X_i=\begin{cases}1 & \text{如果在这 } n \text{ 张票券中,至少有一张是第 } i \text{ 型票券}\\ 0 & \text{如果在这 } n \text{ 张票券中,没有一张是第 } i \text{ 型票券}\end{cases}$$

因为

$$E[X_i]=P\{X_i=1\}=1-P\{\text{在这 } n \text{ 张票券中,没有一张是第 } i \text{ 型票券}\}=1-\left(\frac{N-1}{N}\right)^n$$

所以

$$E[X]=E[X_1]+\cdots+E[X_N]=N\left[1-\left(\frac{N-1}{N}\right)^n\right]$$

(b) 令 Y 表示为获得整套票券所需收集的票券数，我们利用与例 2f 中计算负二项随机变量的均值同样的方法来计算 $E[Y]$. 即，定义 $Y_i(i=0,\ 1,\ \cdots,\ N-1)$为已经得到 i 种不同类型票券后，为获得另一种不同的类型所需收集的票券数. 注意到

$$Y=Y_0+Y_1+\cdots+Y_{N-1}$$

当已收集到 i 种不同类型的票券后，得到与前面不同类型的一张新票券的概率为 $N-i/N$. 因此

$$P\{Y_i=k\}=\frac{N-i}{N}\left(\frac{i}{N}\right)^{k-1}\qquad k\geqslant 1$$

311

换句话说，Y_i 是以$(N-i)/N$ 为参数的几何随机变量. 于是

$$E[Y_i]=\frac{N}{N-i}$$

从而

$$E[Y]=1+\frac{N}{N-1}+\frac{N}{N-2}+\cdots+\frac{N}{1}=N\left[1+\cdots+\frac{1}{N-1}+\frac{1}{N}\right]$$ ■

例 2k 10 个猎人正等着野鸭飞过来，当一群野鸭飞过头顶时，他们同时开了枪，但他们每个人都是随机地、彼此独立地选择自己的目标. 如果每个猎人独立地射中其目标的概率均为 p，试求当 10 只野鸭飞来时，没有被击中而飞走的野鸭数的期望值.

解 设

$$X_i=\begin{cases}1 & \text{第 } i \text{ 只野鸭未被击中而飞走} \\ 0 & \text{第 } i \text{ 只野鸭被击中}\end{cases} \quad i=1,2,\cdots,10$$

飞走的野鸭的期望值可表示为

$$E[X_1+\cdots+X_{10}]=E[X_1]+\cdots+E[X_{10}]$$

为了计算 $E[X_i]=P\{X_i=1\}$，注意到每个猎人独立地射中第 i 只野鸭的概率为$\frac{p}{10}$，故

$$P\{X_i=1\}=\left(1-\frac{p}{10}\right)^{10}$$

因此

$$E[X]=10\left(1-\frac{p}{10}\right)^{10}$$

■

例 2l(游程数的数学期望)　假设由 n 个“1”和 m 个“0”组成的序列随机地排列在一起，共有$(n+m)!\ /n!\ m!$ 种可能的排列，每种排列都是等可能的. 任何连续的“1”排在一起称为一个游程，例如，$n=6$，$m=4$，且排列为 1，1，1，0，1，1，0，0，1，0，那么有 3 个由“1”所组成的游程，下面我们求所有这些排列中平均所含有的游程数. 为此，令

$$I_i=\begin{cases}1 & \text{如果由“1”所组成的一个游程从第 } i \text{ 个位置开始} \\ 0 & \text{其他}\end{cases}$$

因此，由“1”所组成的游程数 $R(1)$可表示为

$$R(1)=\sum_{i=1}^{n+m} I_i$$

因此

312

$$E[R(1)]=\sum_{i=1}^{n+m} E[I_i]$$

由于

$$E[I_1]=P\{\text{“1”在第 1 个位置}\}=\frac{n}{n+m}$$

且对 $1<i\leqslant n+m$，

$$E[I_i]=P\{\text{“0”在第 } i-1 \text{ 个位置，“1”在第 } i \text{ 个位置}\}=\frac{m}{n+m}\,\frac{n}{n+m-1}$$

因此

$$E[R(1)]=\frac{n}{n+m}+(n+m-1)\frac{nm}{(n+m)(n+m-1)}$$

类似地，由“0”所组成的游程数的数学期望为

$$E[R(0)]=\frac{m}{n+m}+\frac{nm}{n+m}$$

所有两种类型的游程数的期望值为

$$E[R(0)+R(1)]=1+\frac{2nm}{n+m}$$

■

例 2m　有 $n+m$ 条指令排列在一起，其中 n 条为特殊指令，m 条为普通指令，假定$(n+m)!$ 种排列方式是等可能的，X 表示第一条特殊指令在排列中所处的位置，求 $E[X]$.

解 因为前 $k-1$ 条指令等可能地为 $\binom{n+m}{k-1}$ 种情况中的任何一种，所以前 $k-1$ 条指令全是普通指令的概率为 $\dfrac{\binom{m}{k-1}}{\binom{n+m}{k-1}}$. 因此，如果前 $k-1$ 条指令全是普通指令，而第 k 条指令是特殊指令，则 $X=k$，故可得

$$P\{X=k\}=\frac{\binom{m}{k-1}}{\binom{n+m}{k-1}}\frac{n}{n+m-k+1},\ k\geqslant 1 \tag{2-3}$$

然而，与其直接运用前面的概率去求，不如赋予 X 一个随机变量和的形式然后去求 $E[X]$. 为此，首先用 o_1，o_2，…，o_m 来表示普通指令，对于 $i=1, 2, \cdots, m$，令

313

$$X_i=\begin{cases}1 & \text{如果 } o_i \text{ 出现在所有 } n \text{ 个特殊指令之前}\\ 0 & \text{其他}\end{cases}$$

由于第一条特殊指令所处的位置比那些所有排在特殊指令之前的普通指令数大 1，故 $X=1+\sum_{i=1}^{m}X_i$

于是

$$E[X]=1+\sum_{i=1}^{m}E[X_i]$$

然而，如果 o_i 排在所有 n 条特殊指令之前，X_i 就等于 1；又因为 o_i 与 n 条特殊指令组成的 $n+1$ 条指令在排列中所处的位置是等可能的，故

$$E[X_i]=P\{X_i=1\}=\frac{1}{n+1}$$

因此

$$E[X]=1+\frac{m}{n+1}$$

从上面的结果可知，一副洗好的纸牌平均需要翻动 $1+\frac{39}{14}=3.786$ 张纸牌才会出现黑桃，平均需要翻动 $1+\frac{48}{5}=10.6$ 张纸牌才能出现一张“A”. 它还有其他方面的应用.

在计算机科学应用方面，假设有 m 条记录(记为 r_1，r_2，…，r_m)分布在 $n+m$ 个位置上，每个位置至多一条记录. 当 m 比 n 小得多时，如何分布这 m 条记录使得当需要它们时能尽可能快地取出？一个有效的方法是使用*散列函数*(hashing function)h 把记录映射到某一位置上. 好的散列函数能使映射的记录值为一个独立随机变量序列，每一个等可能地分布在 $n+m$ 个位置的任何一个位置. 使用散列函数 h，能使记录 r_1 安放在位置 $h(r_1)$上；假设位置 $h(r_2)$未被占用，记录 r_2 安放在位置 $h(r_2)$上，假设位置 $h(r_2)$已被占用，则记录 r_2 后移安放在位置 $h(r_2)+1$ 上. 通常，一旦记录 r_1，r_2，…，r_k 均已被安放在各自位置上，记录 r_{k+1} 将被安放在位置 $h(r_{k+1})+j$ 上，其中 j 是满足位置 $h(r_{k+1})+j$ 未被占用的最小的非负整数. 换言之，记录 r_{k+1} 被放到位置 $h(r_{k+1})$上；如果位置 $h(r_{k+1})$被占用，将继续移动到后面的位置直至找到空位置为

止.(全部移动按 mod$n+m$ 进行,因此,如果一个记录发现位置 $n+m$ 被占用,则移到位置1.)

如果出现一个记录移到某个位置上时,而这个位置已被占用,就称为发生冲突. 已知 m 条记录已放好位置,而未发生冲突[因为所有位置 $h(r_1)$,…,$h(r_m)$是各不相同的],假定记录的散列值是等可能地放到 $m+n$ 个位置的任何一个位置,求安排一个新的记录所引起的冲突的期望值.

314

首先,由于 m 条记录已被放好而未发生任何冲突,由此得出它们最初全部被放在不同的位置上. 由于所有位置是等可能的,这意味着$\binom{n+m}{m}$个子集中的每一个子集的 m 个位置等可能地是这些记录的位置集. 现在考虑新的记录. 它的散列值等可能地是 $n+m$ 个值中的任何一个,且安放它的位置所引起的冲突数等于不空的位置数,即从它的散列值处开始连续向右移动直到找到一个空位置为止所搜索的次数. 但是,通过把 n 个空的位置看作特殊指令的位置,m 个不空的位置看作普通指令的位置,可以得到冲突数,记作 N,它与 n 条特殊指令和 m 条普通指令的随机排序中在所有的特殊指令之前的普通指令数有相同的分布. 因此,$X\equiv N+1$ 服从式(2-3)给出的分布,故

$$E[N]=E[X]-1=\frac{m}{n+1}$$ ■

例 2n(平面上的随机运动) 考虑一个粒子,它开始位于平面上的某点,假定它以固定步长沿着完全随机的方向进行一系列运动. 特别地,假定它每步移动一个单位的距离,且每次移动时的方向角(与水平线的夹角)为$(0, 2\pi)$上的均匀分布(见图7-1),试计算 n 次运动后与原点间距离的平方的期望值.

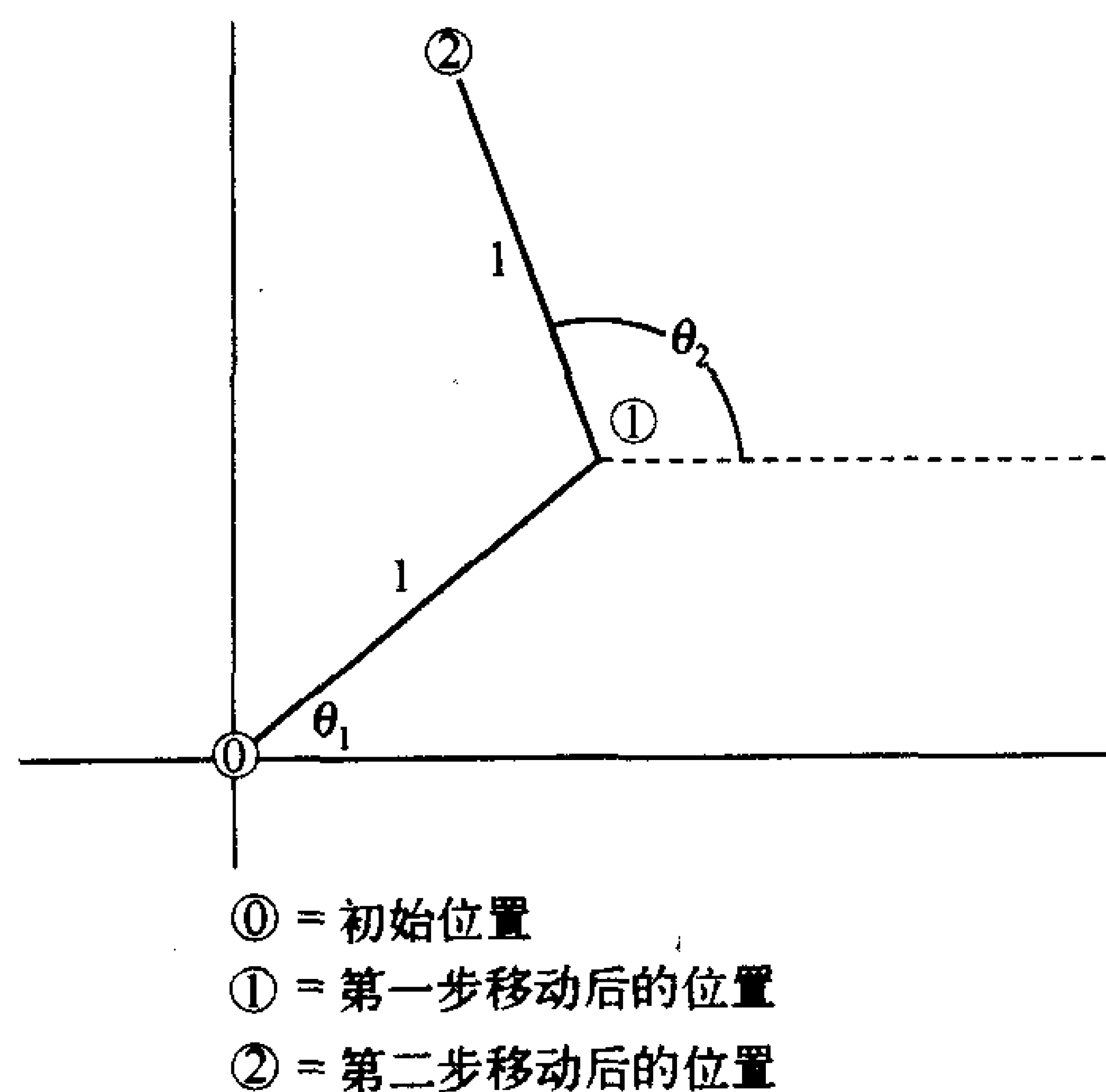


图 7-1

解 令(X_i, Y_i)表示第 $i(i=1, 2, \cdots, n)$步移动后位置的变化的直角坐标,则有

$$X_i=\cos\theta_i, \qquad Y_i=\sin\theta_i$$

其中 $\theta_i(i=1, 2, \cdots, n)$是相互独立的且服从$(0, 2\pi)$上的均匀分布的随机变量. 因为经过 n 步移动后粒子的直角坐标为$\left(\sum_{i=1}^{n}X_i, \sum_{i=1}^{n}Y_i\right)$,我们用 D^2 表示粒子与原点的距离的平方,则

$$\begin{aligned}D^2 &= \left(\sum_{i=1}^{n}X_i\right)^2+\left(\sum_{i=1}^{n}Y_i\right)^2\\&=\sum_{i=1}^{n}(X_i^2+Y_i^2)+\sum_{i\neq j}\sum(X_iX_j+Y_iY_j)\\&=n+\sum_{i\neq j}\sum(\cos\theta_i\cos\theta_j+\sin\theta_i\sin\theta_j)\end{aligned}$$

这里 $\cos\theta_i^2+\sin\theta_i^2=1$. 取数学期望并利用随机变量 θ_i 和 $\theta_j(i\neq j)$的独立性得到

$$2\pi E[\cos\theta_i]=\int_0^{2\pi}\cos u\,du=\sin 2\pi-\sin 0=0$$

$$2\pi E[\sin\theta_i] = \int_0^{2\pi} \sin u \, du = \cos 0 - \cos 2\pi = 0$$

我们得到

$$E[D^2] = n$$

■

例 2o（快速排序算法分析） 假设有 n 个不同的值 $x_1, x_2, \cdots, x_n$，我们需要把它们按照递增顺序排列，或如通常所说的将它们进行排序（sort）。一个有效的办法是采用快速排序算法，基本方法如下：当 $n=2$ 时，将两个值进行比较，然后按适当次序排列。当 $n>2$ 时，则从中随机地选择一个元素，比如说 x_i，将其他值与 x_i 进行比较。把比 x_i 小的数放在其左边的括号中，比 x_i 大的数放在其右边的括号中，然后对括号中的值重复以上步骤直到所有的值被排序为止。例如，假设我们需要将下列 10 个不同的数字进行排序：

315～316

$$5,\ 9,\ 3,\ 10,\ 11,\ 14,\ 8,\ 4,\ 17,\ 6$$

首先随机任选其中一个数（每个数被选到的概率均为 $\frac{1}{10}$），不妨设 10 被选到，我们将其他数与 10 比较，把比 10 小的数放在其左边的括号中，比 10 大的数放在其右边的括号中，得

$$\{5,\ 9,\ 3,\ 8,\ 4,\ 6\},\ 10,\ \{11,\ 14,\ 17\}$$

下面我们看哪一个括号中的数多于一个，首先看左边，从中随机选出一个数，假设为 6，然后将左边括号中的数与 6 比较，把比 6 小的数放在其左边的括号中，比 6 大的数放在其右边的括号中，得

$$\{5,\ 3,\ 4\},\ 6,\ \{9,\ 8\},\ 10,\ \{11,\ 14,\ 17\}$$

看最左边括号中的数，假设随机选出 4，重复上面的步骤，则排列后为

$$\{3\},\ 4,\ \{5\},\ 6,\ \{9,\ 8\},\ 10,\ \{11,\ 14,\ 17\}$$

一直重复以上步骤直到所有括号中的数均为一个。

如果令 X 表示对 n 个不同的数采用快速排序算法进行排序所需比较的次数，则 $E[X]$ 是这种算法的效率的一个度量。为计算 $E[X]$，我们可把 X 表示成如下的随机变量和的形式：首先给参加排序的值一个记号，将值最小的数记为 1，次小的数记为 2，等等。对于 $1 \leqslant i < j \leqslant n$，令

$$I(i,j) = \begin{cases} 1 & \text{如果 } i \text{ 和 } j \text{ 曾经进行过比较} \\ 0 & \text{否则} \end{cases}$$

则

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(i,j)$$

这蕴涵着

$$E[X] = E\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(i,j)\right] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[I(i,j)] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P\{i \text{ 和 } j \text{ 曾经进行过比较}\}$$

为计算 i 和 j 曾经比较过次序的概率，注意到 $i, i+1, \cdots, j-1, j$ 最初处于同一个括号中（因为所有的值最初在同一个括号中），而且如果第一次被选中的数不在 i 和 j 之间则它们一直处于同一个括号中。例如，若选中的比较数比 j 大，则元素 $i, i+1, \cdots, j-1, j$ 将处于被选中的数左边的括号中，如果被选中的数比 i 小，则它们将处于其右边的括号中，于是 $i, i+1, \cdots, j-1, j$ 始终处于同一个括号中直到它们中间的一个数被选中去作比较，此时 i 和 j 之间的其他数都将和这个数比较。如果这个数不是 i 或 j，则 i 和 j 将分别处于这个数的左、右括号中且二者不会再进行比较。另一方面，如果被选中的元素为 i 或 j，则 i 和 j 将进行直接比较。现在，

317

假设选中的数在 i 和 j 之间，且这 $j-i+1$ 个数中的任何一个均等可能地被选中，选中的元素为 i 或 j 的概率为 $2/(j-i+1)$. 因此可得

$$P\{i \text{ 和 } j \text{ 曾经进行过比较}\}=\frac{2}{j-i+1}$$

且

$$E[X] = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{2}{j-i+1}$$

为了在 n 较大时对 $E[X]$的大小进行粗略的近似，我们可以用积分来代替求和，即

$$\sum_{j=i+1}^{n}\frac{2}{j-i+1} \approx \int_{i+1}^{n}\frac{2}{x-i+1}\mathrm{d}x = 2\log(x-i+1)\Big|_{i+1}^{n}$$
$$= 2\log(n-i+1) - 2\log(2) \approx 2\log(n-i+1)$$

因此

$$E[X] \approx \sum_{i=1}^{n-1}2\log(n-i+1) \approx 2\int_{1}^{n-1}\log(n-x+1)\mathrm{d}x$$
$$= 2\int_{2}^{n}\log(y)\mathrm{d}y = 2(y\log(y)-y)\Big|_{2}^{n} \approx 2n\log(n)$$

由此可以看出当 n 很大时，利用快速排序算法把 n 个数排列好平均需要 $2n\log(n)$次比较. ■

例 2p（事件的并的概率） 设 $A_1, A_2, \cdots, A_n$ 表示一系列事件，并定义示性变量 $X_i(i=1, \cdots, n)$为

$$X_i=\begin{cases}1 & \text{如果 } A_i \text{ 发生}\\ 0 & \text{否则}\end{cases}$$

318

现在，注意到

$$1-\prod_{i=1}^{n}(1-X_i) = \begin{cases}1 & \text{如果 } \bigcup A_i \text{ 发生}\\ 0 & \text{否则}\end{cases}$$

因此

$$E\Big[1-\prod_{i=1}^{n}(1-X_i)\Big]= P\Big(\bigcup_{i=1}^{n} A_i\Big)$$

把上式左边展开可得

$$P\Big(\bigcup_{i=1}^{n} A_i\Big)= E\Big[\sum_{i=1}^{n}X_i - \sum_{i<j}\sum X_iX_j + \sum\sum_{i<j<k}\sum X_iX_jX_k - \cdots + (-1)^{n+1}X_1\cdots X_n\Big] \tag{2-4}$$

然而，由于

$$X_{i_1}X_{i_2}\cdots X_{i_k}=\begin{cases}1 & \text{如果 } A_{i_1}A_{i_2}\cdots A_{i_k} \text{ 发生}\\ 0 & \text{否则}\end{cases}$$

因此

$$E[X_{i_1}\cdots X_{i_k}]=P(A_{i_1}\cdots A_{i_k})$$

式(2-4)就是著名的事件的并的概率公式.

$$P(\bigcup A_i)$$
$$= \sum_i P(A_i) - \sum_{i<j}\sum P(A_iA_j) + \sum\sum_{i<j<k}\sum P(A_iA_jA_k) - \cdots + (-1)^{n-1}P(A_1\cdots A_n)$$ ■

当我们处理无穷多个随机变量 $X_i(i\geqslant 1)$时，即使每个随机变量都有有穷的数学期望，

$$E\left[\sum_{i=1}^{\infty} X_i\right]=\sum_{i=1}^{\infty} E[X_i] \tag{2-5}$$

也未必成立. 为了确定式(2-5)何时成立，注意到$\sum_{i=1}^{\infty} X_i=\lim_{n\to\infty}\sum_{i=1}^{n} X_i$，于是

$$E\left[\sum_{i=1}^{\infty} X_i\right]=E\left[\lim_{n\to\infty}\sum_{i=1}^{n} X_i\right]\overset{?}{=}\lim_{n\to\infty}E\left[\sum_{i=1}^{n} X_i\right]=\lim_{n\to\infty}\sum_{i=1}^{n}E[X_i]=\sum_{i=1}^{\infty}E[X_i] \tag{2-6}$$

319

因此，当式(2-6)中的期望和极限运算可以交换次序时，式(2-5)成立. 尽管一般情况下这种交换是不满足的，但对以下两种特例情况，式(2-5)成立：

1. 随机变量 X_i 都是非负的(即 $P\{X_i\geqslant 0\}=1$,对一切 i).
2. $\sum_{i=1}^{\infty} E[|X_i|]<\infty$.

例 2q 对于非负取整数值的随机变量 X，如果对每个 $i\geqslant 1$，我们定义

$$X_i=\begin{cases}1 & \text{如果 } X\geqslant i\\ 0 & \text{如果 } X<i\end{cases}$$

则

$$\sum_{i=1}^{\infty} X_i=\sum_{i=1}^{X} X_i+\sum_{i=X+1}^{\infty} X_i=\sum_{i=1}^{X}1+\sum_{i=X+1}^{\infty}0=X$$

由于 X_i 都是非负的，因此

$$E[X]=\sum_{i=1}^{\infty}E(X_i)=\sum_{i=1}^{\infty}P\{X\geqslant i\} \tag{2-7}$$

这是一个有用的恒等式. ■

例 2r 假设有 n 个元素，记为 $1, 2, \cdots, n$，按照一定的顺序存储在计算机中，每单位时间计算机实施的命令需要这些元素中的一个，每个元素 i 被选中的概率为 $P(i)$，且与上次是否选中无关. 假设概率 $P(i)$ 已知，且 $\sum_i P(i)=1$，问什么样的排序才能使计算机所选取的元素所处的位置期望值最小?

解 假设元素被排放使得 $P(1)\geqslant P(2)\geqslant\cdots\geqslant P(n)$. 为了证明 $1, 2, \cdots, n$ 是最优的排序，令 X 表示所需要的元素所处的位置，在任一种排序 $O=i_1, i_2, \cdots, i_n$ 下，

$$P_O\{X\geqslant k\}=\sum_{j=k}^{n}P(i_j)\geqslant\sum_{j=k}^{n}P(j)=P_{1,2,\cdots,n}\{X\geqslant k\}$$

320

对所有 k 求和并运用式(2-7)得

$$E_O[X]\geqslant E_{1,2,\cdots,n}[X]$$

因此按照元素被选中的概率降序排列能使选取元素的位置期望值最小. ■

例 2s 在例 2m 中看到，如果用 X 表示 n 条特殊指令和 m 条普通指令随机排列时第一条特殊指令所处的位置，则

$$E[X]=\frac{n+m+1}{n+1}$$

由于当前 $k-1$ 条指令均是普通指令时 $X\geqslant k$，

$$P\{X\geqslant k\}=\frac{\binom{m}{k-1}}{\binom{n+m}{k-1}}$$

因此，由式(2-7)得到

$$\sum_{k=1}^{n}\frac{\binom{m}{k-1}}{\binom{n+m}{k-1}}=\frac{n+m+1}{n+1}$$

■

*7.2.1 用概率方法得到数学期望的界

概率方法是通过对某一集合上的元素赋予概率来分析这个集合中元素的性质，然后研究所选元素与这些概率的对应. 这种方法在第 3 章例 4k 中已有介绍，那里我们用它来证明满足一定的性质的集合元素的存在性. 本小节我们说明它有时可以用来求复杂函数界的问题.

设 f 为定义在一个有限集$\mathcal{S}$的元素上的函数，且假设我们对

$$m=\max_{s\in\mathcal{S}}f(s)$$

感兴趣. m 的一个有用的下界常采用下面的方法来得到：令 S 为集合$\mathcal{S}$上的随机元素，$f(S)$的期望是可以计算的，注意到 $m\geqslant f(S)$可得

$$m\geqslant E[f(S)]$$

321 如果 $f(S)$不是一个常数随机变量，严格不等式是成立的. 即 $E[f(S)]$是函数最大值的下界.

例 2t(比赛中哈密顿路的最大数目) 某循环赛有 $n(n>2)$个选手参加，所有$\binom{n}{2}$场可能的比赛中每对选手只进行一次比赛. 假定选手被记为 1，2，…，n，排列 i_1，i_2，…，i_n 称为哈密顿路(Hamiltonian path)，如果 i_1 胜 i_2，i_2 胜 i_3，…，i_{n-1}胜 i_n. 一个有趣的问题是确定哈密顿路的最大数目.

作为说明，假设有 3 名选手参加比赛，如果他们中有一人胜了两次，则只有一个哈密顿路(例如，选手 1 胜了两次，且选手 2 胜了选手 3，则唯一的哈密顿路为 1，2，3)；另一方面，如果每个选手都胜一次，则将有 3 个哈密顿路(例如，选手 1 胜选手 2，选手 2 胜选手 3，选手 3 胜选手 1，则 1，2，3，2，3，1 以及 3，1，2 都是哈密顿路). 因此当 $n=3$ 时，最多有 3 个哈密顿路.

我们将证明有一种比赛的结果使得哈密顿路的数目大于 $n!/2^{n-1}$. 首先将所有的比赛结果规定为$\binom{n}{2}$场比赛的每一个结果，令$\mathcal{S}$表示所有 $2^{\binom{n}{2}}$ 种可能的结果构成的集合. 用 $f(s)$表示结果 $s\in\mathcal{S}$所对应的哈密顿路的数目，我们将证明

$$\max_{s}f(s)\geqslant\frac{n!}{2^{n-1}}$$

为此，随机地选择所得的结果 S，使所有$\binom{n}{2}$场比赛结果彼此独立，且每个选手等可能地胜每场比赛. 为计算结果 S 的哈密顿路的数目的数学期望$E[f(S)]$，对于 $n!$ 种排列，令 $i=1，2，\cdots，n!$，定义

$$X_i=\begin{cases}1 & \text{如果第 } i \text{ 种排列是哈密顿路}\\ 0 & \text{否则}\end{cases}$$

由于

$$f(S)=\sum_{i}X_i$$

因此

$$E[f(S)] = \sum_i E[X_i]$$

根据比赛结果独立性的假定，任意一种指定的排列是一个哈密顿路的概率是$(1/2)^{n-1}$，则有

$$E[X_i]=P\{X_i=1\}=(1/2)^{n-1}$$

322

因此

$$E[f(S)]=n!\ (1/2)^{n-1}$$

由于$f(S)$不是一个常数随机变量，故从上面等式可知存在一个比赛结果含有多于$n!\ /2^{n-1}$个哈密顿路. ■

例 2u 一片小树林有 52 棵树环形排列在一起，假设有 15 只松鼠生活在这些树上，证明存在 7 棵相连的树上至少住有 3 只松鼠.

解 设一棵树按顺时针方向与其相邻的 6 棵树组成一个邻域. 对于 15 只松鼠生活场所的任何一种选择，我们需要证明存在一棵树，在其邻域中至少有 3 只松鼠. 为此，随机地选一棵树，用 X 表示它所在邻域中的松鼠的数目. 为计算 $E[X]$，对于 15 只松鼠中的任何一个，令 $i=1, 2, \cdots, 15$，定义

$$X_i=\begin{cases}1 & \text{如果松鼠 } i \text{ 生活在随机选择的树的邻域内}\\ 0 & \text{否则}\end{cases}$$

因为

$$X = \sum_{i=1}^{15} X_i$$

所以

$$E[X] = \sum_{i=1}^{15} E[X_i]$$

然而，假如随机选择的树是 7 棵树中的任意一棵，松鼠 i 生活在该树的顺时针方向的 6 棵相邻的树上，则 $X_i=1$,故

$$E[X_i]=P\{X_i=1\}=\frac{7}{52}$$

$$E[X]=\frac{105}{52}>2$$

这就证明存在一棵树，在它的邻域中生活的松鼠多于 2 只. ■

323

*7.2.2 最大－最小恒等式

我们首先介绍一个数集上的最大值与其子集上的最小值之间的关系的恒等式：

命题 2.2 对任意数集 x_i，$i=1, 2, \cdots, n$，

$$\max_i x_i = \sum_i x_i - \sum_{i<j}\min(x_i,x_j) + \sum_{i<j<k}\min(x_i,x_j,x_k) + \cdots + (-1)^{n+1}\min(x_1,\cdots,x_n)$$

证明 我们将用概率方法证明这个结果. 首先假设所有的 x_i 均在区间$[0, 1]$上，令 U 表示区间$(0, 1)$上的均匀随机变量，且由 $A_i=\{U<x_i\}$定义事件 A_i，$i=1, \cdots, n$，即 A_i 表示均匀随机变量小于 x_i 的事件. 如果 U 至少小于 x_i 中的一个，则至少有一个 A_i 发生，可得

$$\bigcup_i A_i=\{U<\max_i x_i\}$$

因此，

$$P\Big(\bigcup_i A_i\Big)=P\{U<\max_i x_i\}=\max_i x_i$$

又

$$P(A_i)=P\{U<x_i\}=x_i$$

另外，如果 U 比所有的 x_{i_1}，…，x_{i_r} 都小，则所有的事件 A_{i_1}，…，A_{i_r} 均会发生，可以看到这些事件的交为

$$A_{i_1}\cdots A_{i_r}=\{U<\min_{j=1,\cdots,r} x_{i_j}\}$$

可得

$$P(A_{i_1}\cdots A_{i_r})=P\{U<\min_{j=1,\cdots,r} x_{i_j}\}=\min_{j=1,\cdots,r} x_{i_j}$$

则由事件的并的概率公式可得

$$P(\bigcup_i A_i)=\sum_i P(A_i)-\sum_{i<j}P(A_iA_j)+\sum_{i<j<k}P(A_iA_jA_k)+\cdots+(-1)^{n+1}P(A_1\cdots A_n)$$

因此命题成立.

当 x_i 为非负数但并不限定在单位区间上时，可找到一个数 c 使得所有的 x_i 都小于 c，则恒等式对于 $y_i=x_i/c$ 成立，所得结果两边同乘以 c 可得命题成立. 当 x_i 可以为负数时，可找到一个 b 使得 $x_i+b>0$ 对所有的 i 成立，则

324

$$\max_i(x_i+b)=\sum_i(x_i+b)-\sum_{i<j}\min(x_i+b,x_j+b)+\cdots+(-1)^{n+1}\min(x_1+b,\cdots,x_n+b)$$

令

$$M=\sum_i x_i-\sum_{i<j}\min(x_i,x_j)+\cdots+(-1)^{n+1}\min(x_1\cdots,x_n)$$

可重写上面的恒等式为

$$\max_i x_i+b=M+b\left(n-\binom{n}{2}+\cdots+(-1)^{n+1}\binom{n}{n}\right)$$

然而

$$0=(1-1)^n=1-n+\binom{n}{2}+\cdots+(-1)^n\binom{n}{n}$$

上面两个等式表明

$$\max_i x_i=M$$

从而命题得证. ■

由命题 2.2 可知，对任意的随机变量 X_1，…，X_n 有

$$\max_i X_i=\sum_i X_i-\sum_{i<j}\min(X_i,X_j)+\cdots+(-1)^{n+1}\min(X_1,\cdots,X_n)$$

对上式两边取数学期望可得最大值与部分最小值之间的关系：

$$E[\max_i X_i]=\sum_i E[X_i]-\sum_{i<j}E[\min(X_i,X_j)]+\cdots+(-1)^{n+1}E[\min(X_1,\cdots,X_n)] \quad (2\text{-}8)$$

例 2v（不等概率的票券收集） 假设有 n 种不同类型的票券，每次从中任意抽取一张，且与前面收集的票券相互独立，抽到第 i 种票券的概率为 p_i，$\sum_{i=1}^{n}p_i=1$. 为了得到每种票券至少一张的一个完全集，求需要收集的票券数的期望值.

解 令 X_i 表示得到第 i 种票券需要收集的票券数，则有

325

$$X=\max_{i=1,\cdots,n} X_i$$

因为得到的每一种新票券是第 i 种票券的概率为 p_i，可以看到 X_i 是服从参数为 p_i 的几何随机变量. 又因为 X_i 与 X_j 的最小值是得到类型 i 或类型 j 所需要收集的票券数，可以看到对 $i\neq j$，$\min(X_i, X_j)$是服从参数为 p_i+p_j 的几何随机变量. 同理 $\min(x_i, x_j, x_k)$是服从参数为 $p_i+p_j+p_k$ 的几何随机变量，等等. 因此，由式(2-8)可得

$$E[X]=\sum_i \frac{1}{p_i}-\sum_{i<j}\frac{1}{p_i+p_j}+\sum_{i<j<k}\frac{1}{p_i+p_j+p_k}+\cdots+(-1)^{n+1}\frac{1}{p_1+\cdots+p_n}$$

又

$$\int_0^{\infty} e^{-px}\,dx=\frac{1}{p}$$

利用恒等式

$$1-\prod_{i=1}^{n}(1-e^{-p_i x})=\sum_i e^{-p_i x}-\sum_{i<j}e^{-(p_i+p_j)x}+\cdots+(-1)^{n+1}e^{-(p_1+\cdots+p_n)x}$$

对上式两边积分得

$$E[X]=\int_0^{\infty}\left(1-\prod_{i=1}^{n}(1-e^{-p_i x})\right)dx$$

这是一个更有用的计算公式. ■

例 2w 假设一副牌共 n 张，其中第 i 种类型的牌有 $n_i(i=1,\cdots,r)$张，$\sum_{i=1}^{r}n_i=n$，洗牌使得所有 $n!$ 种排列都是等可能的，每次翻开一张牌.

(a) 试确定平均需要翻开多少张牌才能使所有类型的牌都至少出现一次.

(b) 试确定平均需要翻开多少张牌才能使其中一种类型的牌全部出现.

解 假如令 X_i 表示直至第 i 种牌出现需要翻开的张数，则使所有类型的牌都至少出现一次需要的翻开次数 X 可表示为

$$X=\max_{i=1,\cdots,r}X_i$$

由于第 i 种牌有 n_i 张，由例 2m 的结论可知为获得第 i 种牌平均需要的翻动次数为

$$E[X_i]=1+\frac{n-n_i}{n_i+1}=\frac{n+1}{n_i+1}$$

又对于 $i\neq j$，$\min(X_i, X_j)$是获得第 i 种或第 j 种牌所需翻动的次数，由例 2m 可得

$$E[\min(X_i,X_j)]=\frac{n+1}{n_i+n_j+1}$$

326

其他的极小值的数学期望同理可得，应用式(2-8)可得

$$E[X]=\sum_i\frac{n+1}{n_i+1}-\sum_{i<j}\frac{n+1}{n_i+n_j+1}+\cdots+(-1)^{r+1}\frac{n+1}{n_1+\cdots+n_r+1}$$

又

$$\int_0^1 x^m\,dx=\frac{1}{m+1}$$

应用等式

$$1-\prod_{i=1}^{r}(1-x^{n_i})=\sum_i x^{n_i}-\sum_{i<j}x^{n_i+n_j}+\cdots+(-1)^{r+1}x^{n_1+\cdots+n_r}$$

对上式积分可得

$$E[X]=(n+1)\int_0^1\left(1-\prod_{i=1}^{r}(1-x^{n_i})\right)dx \tag{2-9}$$

(b)为确定在得到至少一张第 r 类型的牌的一个完全集之前需要翻开的牌数的数学期望，注意到当余下的牌中不再包含每类牌中的至少一张牌的话，就已经构成了一个完全集. 由于一副牌从后到前的排列是等可能地是 $n!$ 种排列中的任何一种，因此，一副牌从后翻到前，得到每个类型至少一张牌需要翻开的牌数的数学期望就是 $E[X]$. 从而，当得到至少有一张第 r 类型的牌的一个完全集时，余下的牌数的数学期望为 $E[X]-1$，这就表明为得到某种类型的所有牌需要翻开的牌数的数学期望为 $n-E[X]+1$，这里 $E[X]$由式(2-9)给出. ■

7.3 协方差、和的方差与相关系数

首先我们给出下面的命题，即独立随机变量的乘积的数学期望等于它们的数学期望的乘积.

327

命题 3.1 如果 X 和 Y 为独立随机变量，则对任意的函数 h 和 g 有

$$E[g(X)h(Y)]=E[g(X)]E[h(Y)]$$

证明 假设随机变量 X 和 Y 有连续的联合密度函数 $f(x, y)$，则

$$\begin{aligned}E[g(X)h(Y)] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y)f(x,y)\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y)f_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{+\infty} h(y)f_Y(y)\mathrm{d}y\int_{-\infty}^{+\infty} g(x)f_X(x)\mathrm{d}x = E[h(Y)]E[g(X)]\end{aligned}$$

对于离散情形，证法类似. ■

如同数学期望和方差描述单个随机变量的性质一样，对于两个随机变量之间的性质我们用两个变量的协方差来描述.

定义 随机变量 X 和 Y 的协方差 $\mathrm{Cov}(X, Y)$定义为

$$\mathrm{Cov}(X,Y)=E[(X-E[X])(Y-E[Y])]$$

对上式右边展开可得

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E[XY-E[X]Y-XE[Y]+E[X]E[Y]] \\ &= E[XY]-E[X]E[Y]-E[X]E[Y]+E[X]E[Y] \\ &= E[XY]-E[X]E[Y]\end{aligned}$$

由命题 3.1 可知，如果 X 和 Y 相互独立，则 $\mathrm{Cov}(X, Y)=0$. 然而此逆命题不成立，两个非独立的随机变量 X 和 Y 的协方差等于零的一个简单例子如下：设 X 是一个随机变量，使得

$$P\{X=0\}=P\{X=1\}=P\{X=-1\}=\frac{1}{3}$$

328

定义

$$Y=\begin{cases}0 & \text{如果 } X\neq 0\\ 1 & \text{如果 } X=0\end{cases}$$

则 $XY=0$，因此 $E[XY]=0$. 另外，$E[X]=0$，于是

$$\mathrm{Cov}(X,Y)=E[XY]-E[X]E[Y]=0$$

然而，随机变量 X 和 Y 显然不独立.

下面给出协方差的几个性质.

命题 3.2

(i) $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$.

(ii) $\mathrm{Cov}(X,X)=\mathrm{Var}(X)$.

(iii) $\mathrm{Cov}(aX,Y)=a\mathrm{Cov}(X,Y)$.

(iv) $\mathrm{Cov}\left(\sum_{i=1}^{n}X_i,\sum_{j=1}^{m}Y_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{Cov}(X_i,Y_j)$.

证明 (i)和(ii)由协方差的定义立即可得，(iii)留作习题. 下面证明(iv)，即证明协方差的可加性，首先令 $\mu_i=E[X_i]$，$v_j=E[Y_j]$，则有

$$E\left[\sum_{i=1}^{n}X_i\right]=\sum_{i=1}^{n}\mu_i,\quad E\left[\sum_{j=1}^{m}Y_j\right]=\sum_{j=1}^{m}v_j$$

且

$$\begin{aligned}\mathrm{Cov}\left(\sum_{i=1}^{n}X_i,\sum_{j=1}^{m}Y_j\right)&=E\left[\left(\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}\mu_i\right)\left(\sum_{j=1}^{m}Y_j-\sum_{j=1}^{m}v_j\right)\right]\\&=E\left[\sum_{i=1}^{n}(X_i-\mu_i)\sum_{j=1}^{m}(Y_j-v_j)\right]\\&=E\left[\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i-\mu_i)(Y_j-v_j)\right]\\&=\sum_{i=1}^{n}\sum_{j=1}^{m}E[(X_i-\mu_i)(Y_j-v_j)]\end{aligned}$$

最后一个等式应用了随机变量和的数学期望等于它们各自的数学期望的和. ■ 329

从上面命题3.2的(ii)和(iv)可知，如果令 $Y_j=X_j(j=1,\cdots,n)$，则

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{n}X_i\right)&=\mathrm{Cov}\left(\sum_{i=1}^{n}X_i,\sum_{j=1}^{n}X_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i,Y_j)\\&=\sum_{i=1}^{n}\mathrm{Var}(X_i)+\sum_{i\neq j}\sum\mathrm{Cov}(X_i,X_j)\end{aligned}$$

由于指标 i，$j(i\neq j)$在上面的求和中出现两次，上面的等式等价为

$$\boxed{\mathrm{Var}\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}\mathrm{Var}(X_i)+2\sum_{i<j}\sum\mathrm{Cov}(X_i,X_j)}\qquad(3\text{-}1)$$

如果 X_1，…，X_n 两两相互独立，即对于 $i\neq j$，随机变量 X_i 和 X_j 独立，则式(3-1)变为

$$\mathrm{Var}\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}\mathrm{Var}(X_i)$$

下面给出式(3-1)的一个应用.

例 3a 设 X_1，…，X_n 为独立同分布随机变量，有相同的数学期望 μ 和方差 σ^2，如同例2c中所示，令 $\overline{X}=\dfrac{\sum_{i=1}^{n}X_i}{n}$ 表示样本均值，则变量 $X_i-\overline{X}(i=1,\cdots,n)$ 称为样本偏差(deviation)，随机变量 $S^2=\sum_{i=1}^{n}\dfrac{(X_i-\overline{X})^2}{n-1}$ 称为样本方差(sample variance)，求(a) $\mathrm{Var}(\overline{X})$ 和(b) $E[S^2]$.

解 (a)
$$\begin{aligned}\mathrm{Var}(\overline{X})&=\left(\frac{1}{n}\right)^2\mathrm{Var}\left(\sum_{i=1}^{n}X_i\right)\\&=\left(\frac{1}{n}\right)^2\sum_{i=1}^{n}\mathrm{Var}(X_i)\qquad\text{由独立性}\\&=\frac{\sigma^2}{n}\end{aligned}$$

330

(b) 用下面的代数等式：

$$\begin{aligned}(n-1)S^2 &= \sum_{i=1}^{n}(X_i-\mu+\mu-\overline{X})^2\\ &= \sum_{i=1}^{n}(X_i-\mu)^2+\sum_{i=1}^{n}(\overline{X}-\mu)^2-2(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu)\\ &= \sum_{i=1}^{n}(X_i-\mu)^2+n(\overline{X}-\mu)^2-2(\overline{X}-\mu)n(\overline{X}-\mu)\\ &= \sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\end{aligned}$$

对上式两边取数学期望可得

$$\begin{aligned}(n-1)E[S^2] &= \sum_{i=1}^{n}E[(X_i-\mu)^2]-nE[(\overline{X}-\mu)^2]\\ &= n\sigma^2-n\text{Var}(\overline{X}) = (n-1)\sigma^2\end{aligned}$$

其中最后一个等式用到(a)的结果，第一个等式用到例 2c 的结果 $E[\overline{X}]=\mu$. 再对上式两边同除以 $n-1$，可得样本方差的数学期望等于原分布的方差 σ^2. ■

下面的例子给出了另一种计算二项随机变量方差的方法(在第 4 章中已介绍过一种).

例 3b(二项随机变量的方差) 求参数为 n 和 p 的二项随机变量 X 的方差.

解 由于这种随机变量表示成功率为 p 的 n 次独立试验中成功的次数，可记

$$X=X_1+\cdots+X_n$$

这里 X_i 表示独立的伯努利随机变量，满足

$$X_i=\begin{cases}1 & \text{如果第 } i \text{ 次试验成功}\\ 0 & \text{否则}\end{cases}$$

因此，由式(3-1)可得

331

$$\text{Var}(X)=\text{Var}(X_1)+\cdots+\text{Var}(X_n)$$

但是

$$\begin{aligned}\text{Var}(X_i)&=E[X_i^2]-(E[X_i])^2\\ &=E[X_i]-(E[X_i])^2 \qquad \text{由于 } X_i^2=X_i\\ &=p-p^2\end{aligned}$$

从而

$$\text{Var}(X)=np(1-p)$$ ■

例 3c(配对数目的方差) 在例 2h 中，令 X 表示刚好拿到自己的帽子的人数，求 X 的方差.

解 仍采用例 2h 中对 X 的表示方法，即

$$X=X_1+\cdots+X_N$$

其中

$$X_i=\begin{cases}1 & \text{如果第 } i \text{ 个人拿到自己的帽子}\\ 0 & \text{否则}\end{cases}$$

由式(3-1)得

$$\text{Var}(X) = \sum_{i=1}^{N}\text{Var}(X_i)+2\sum_{i<j}\sum\text{Cov}(X_i,X_j) \tag{3-2}$$

由于 $P\{X_i=1\}=1/N$，由上例可以看到

$$\text{Var}(X_i)=\frac{1}{N}\left(1-\frac{1}{N}\right)=\frac{N-1}{N^2}$$

另外，

$$\mathrm{Cov}(X_i, X_j)=E[X_iX_j]-E[X_i]E[X_j]$$

现在，

$$X_iX_j=\begin{cases}1 & \text{如果第 } i \text{ 个和第 } j \text{ 个人均拿到自己的帽子}\\ 0 & \text{否则}\end{cases}$$

因此

$$E[X_iX_j]=P\{X_i=1, X_j=1\}=P\{X_i=1\}P\{X_j=1 \mid X_i=1\}=\frac{1}{N}\frac{1}{N-1}$$

故

$$\mathrm{Cov}(X_i, X_j)=\frac{1}{N(N-1)}-\left(\frac{1}{N}\right)^2=\frac{1}{N^2(N-1)}$$

332

由式(3-2)得

$$\mathrm{Var}(X)=\frac{N-1}{N}+2\binom{N}{2}\frac{1}{N^2(N-1)}=\frac{N-1}{N}+\frac{1}{N}=1$$

即随机变量 X 的数学期望和方差均为 1. 在某种程度上，这也是意料中的事，在第 2 章例 5m 中，当 N 很大时第 i 个人拿到自己的帽子的概率接近 $e^{-1}/i!$. 即当 N 很大时，配对数目近似服从均值为 1 的泊松分布，又由于泊松分布的均值和方差相等，因此我们得到上面的结果不足为奇. ■

例 3d(负超几何分布) 假设有 $n+m$ 条指令随机排列在一起，其中有 n 条特殊指令. m 条普通指令，且$(n+m)!$ 种可能的排列是等可能的. 如果令 X 表示第 r 条特殊指令在排序中所处的位置，那么称 X 是以(r, n, m)为参数的**负超几何随机变量**(negative hypergeometric random variable). (一个重要的特例是 $r=1$，在例 2m 中已经求出其均值.)为了理解其具体的意义，首先注意这个随机变量和从一个有 n 个特殊球和 m 个普通球的罐中不放回地连续取球，每次从留在罐中的球中等可能地选一球，取到第 r 个特殊球所需要的取球次数有相同的分布. (为了理解它们之间的联系，假定全部 $n+m$ 个球被取出并按其指标排序.)因此，负超几何分布与超几何分布之间的关系和负二项分布与二项分布之间的关系一样，即在一系列独立试验中，结果为成功(抽到特殊球)或失败，二项分布和超几何分布描述的是在一系列固定数目的试验中成功的次数，而负二项分布和负超几何分布描述的是为获得固定数目的成功试验所需要进行的试验次数. (二项分布和超几何分布之间的区别为：二项分布描述的是每次从罐中取球时，要把上一次取的球放回原罐中的情况，而超几何分布描述的是每次取球时，不需要把上次取的球放回原罐中的情况.)

为得到负超几何随机变量 X 的概率质量函数，注意到 $X=k$，如果

(a) 前 $k-1$ 次所抽的球中包括 $r-1$ 个特殊球和 $k-r$ 个普通球，

(b) 第 k 次所取的球为特殊球.

333

因此，

$$P\{X=k\}=\frac{\binom{n}{r-1}\binom{m}{k-r}}{\binom{n+m}{k-1}}\frac{n-r+1}{n+m-k+1}$$

这里不直接用上面的概率质量函数来计算随机变量的期望和方差，而是采用例 2m 中的方法来求，并把 X 表示为随机变量和的形式. 为此，用 o_1，…，o_m 来表示题中的 m 个普通球，且定义随机变量 X_1，…，X_m 为

$$X_i=\begin{cases}1 & \text{如果 } o_i \text{ 在 } r \text{ 个特殊球取走之前被取走}\\ 0 & \text{否则}\end{cases}$$

由于为获得 r 个特殊球需要的取球次数为 r 加上在获得第 r 个特殊球前所取出的普通球的数目，所以有

$$X=r+\sum_{i=1}^{m}X_i \tag{3-3}$$

假设在 $n+1$ 个球中含有 n 个特殊球和普通球 o_i，在这 $n+1$ 个球中，o_i 等可能地被第一次被取出，或第二次被取出，……，或最后一次被取出，因此它在前 r 次被取出的概率为$\frac{r}{n+1}$，故

$$E[X_i]=P\{X_i=1\}=\frac{r}{n+1}$$

意味着

$$E[X]=r+\sum_{i=1}^{m}E[X_i]=r\left(1+\frac{m}{n+1}\right)=\frac{r(n+m+1)}{n+1}$$

为确定 X 的方差，应用式(3-3)可得

$$\mathrm{Var}(X)=\mathrm{Var}\left(\sum_{i=1}^{m}X_i\right)=\sum_{i=1}^{m}\mathrm{Var}(X_i)+\sum_{j}\sum_{i\neq j}\mathrm{Cov}(X_i,X_j)$$

因为 X_i 是参数为 $r/n+1$ 的伯努利随机变量，故

334

$$\mathrm{Var}(X_i)=\frac{r}{n+1}\left(1-\frac{r}{n+1}\right)=\frac{r(n+1-r)}{(n+1)^2}$$

且对于 $i\neq j$，

$$X_iX_j=\begin{cases}1 & \text{如果 } o_i \text{ 和 } o_j \text{ 在 } r \text{ 个特殊球取走之前被取走}\\ 0 & \text{否则}\end{cases}$$

下面考虑 $n+2$ 个球，其中有 o_i，o_j 和 n 个特殊球. 为了使 o_i 和 o_j 在 r 个特殊球取走之前被取走，o_i 和 o_j 必须在取走的 $n+2$ 个球的前 $r+1$ 个之中，因此

$$E[X_iX_j]=P\{X_iX_j=1\}=\frac{\binom{2}{2}\binom{n}{r-1}}{\binom{n+2}{r+1}}=\frac{(r+1)r}{(n+2)(n+1)}$$

对于 $i\neq j$，

$$\mathrm{Cov}(X_i,X_j)=\frac{(r+1)r}{(n+2)(n+1)}-\left(\frac{r}{n+1}\right)^2=\frac{r(n-r+1)}{(n+1)^2(n+2)}$$

所以可得

$$\begin{aligned}\mathrm{Var}(X)&=\frac{mr(n+1-r)}{(n+1)^2}+m(m-1)\frac{r(n-r+1)}{(n+1)^2(n+2)}\\&=\frac{mr(n+1-r)}{(n+1)^2}\left[1+\frac{m-1}{n+2}\right]\\&=\frac{mr(n+1-r)(n+m+1)}{(n+1)^2(n+2)}\end{aligned}$$

■

例 3c（有限人群的抽样问题） 假设有 N 个人，每人对某一特定的事件都有各自的观点，我们把这些观点数量化，以表示此人对该事件的“关心程度”. 用 v_i 表示第 i 人的关心程度，其中 $i=1$，…，N.

假设这些量 v_i 均未知，为收集 N 个人中 n 个人的信息，我们随机地从 N 个人中选取 n 个人，即$\binom{N}{n}$种可能的选取方案是等可能的，选中的 n 个人被调查一些问题并确定他们的感受的

量化值. 如果用 S 来表示 n 个样本值之和, 试求它的均值和方差.

上面的例子有一个很重要的应用, 即用来统计即将举行的选举中广大选民对某一候选人的赞成或反对情况. 如果第 i 个选民持赞成态度则记 $v_i=1$, 否则记 $v_i=0$, 于是 $\bar{v}=\sum_{i=1}^{N} v_i/N$ 就表示所有持赞成态度的选民的比例. 为估计 $\bar{v}$, 随机抽查 n 个选民并记录他们的选票, 持赞成态度的选民所占的比例常用来估计 $\bar{v}$.

335

解 对于第 $i(i=1, \cdots, N)$ 个人, 定义一示性变量 I_i 来表示其是否在被随机抽查之列. 即,

$$I_i=\begin{cases}1 & \text{如果第 } i \text{ 个人在被抽查之列} \\ 0 & \text{否则}\end{cases}$$

则 S 可表示为

$$S=\sum_{i=1}^{N} v_i I_i$$

因此

$$E[S]=\sum_{i=1}^{N} v_i E[I_i]$$

$$\begin{aligned}\operatorname{Var}(S) &= \sum_{i=1}^{N} \operatorname{Var}(v_i I_i)+2 \sum_{i<j}\sum \operatorname{Cov}(v_i I_i, v_j I_j) \\ &= \sum_{i=1}^{N} v_i^2 \operatorname{Var}(I_i)+2 \sum_{i<j}\sum v_i v_j \operatorname{Cov}(I_i, I_j)\end{aligned}$$

由于

$$E[I_i]=\frac{n}{N}$$

$$E[I_i I_j]=\frac{n}{N} \frac{n-1}{N-1}$$

可得

$$\operatorname{Var}(I_i)=\frac{n}{N}\left(1-\frac{n}{N}\right)$$

$$\operatorname{Cov}(I_i, I_j)=\frac{n(n-1)}{N(N-1)}-\left(\frac{n}{N}\right)^2=\frac{-n(N-n)}{N^2(N-1)}$$

因此

$$E[S]=n \sum_{i=1}^{N} \frac{v_i}{N}=n\bar{v}$$

$$\operatorname{Var}(S)=\frac{n}{N}\left(\frac{N-n}{N}\right) \sum_{i=1}^{N} v_i^2-\frac{2n(N-n)}{N^2(N-1)} \sum_{i<j}\sum v_i v_j$$

$\operatorname{Var}(S)$ 的表达式可以用等式 $(v_1+\cdots+v_N)^2=\sum_{i=1}^{N} v_i^2+2 \sum_{i<j}\sum v_i v_j$ 加以简化, 得

336

$$\operatorname{Var}(S)=\frac{n(N-n)}{N-1}\left(\frac{\sum_{i=1}^{N} v_i^2}{N}-\bar{v}^2\right)$$

考虑一种特殊情况, Np 个 v 等于 1, 而其余为 0, 则此时 S 为超几何随机变量, 其均值和方差为

$$E[S]=n\bar{v}=np\left(\text{由于 } \bar{v}=\frac{Np}{N}=p\right)$$

$$\mathrm{Var}(S)=\frac{n(N-n)}{N-1}\left(\frac{Np}{N}-p^2\right)=\frac{n(N-n)}{N-1}p(1-p)$$

变量$\frac{S}{n}$即那些被选中的数值为 1 的比率，满足

$$E\left[\frac{S}{n}\right]=p \quad \mathrm{Var}\left(\frac{S}{n}\right)=\frac{N-n}{n(N-1)}p(1-p)$$

■

随机变量 X 和随机变量 Y 的相关系数$\rho(X, Y)$定义为：当 $\mathrm{Var}(X)\mathrm{Var}(Y)>0$ 时，

$$\rho(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

可以证明

$$-1\leqslant\rho(X,Y)\leqslant 1 \tag{3-4}$$

为证明式(3-4)，假设随机变量 X 和 Y 的方差分别为σ_x^2 和σ_y^2，则

$$0\leqslant\mathrm{Var}\left(\frac{X}{\sigma_x}+\frac{Y}{\sigma_y}\right)=\frac{\mathrm{Var}(X)}{\sigma_x^2}+\frac{\mathrm{Var}(Y)}{\sigma_y^2}+\frac{2\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y}=2[1+\rho(X,Y)]$$

这意味着

337

$$-1\leqslant\rho(X,Y)$$

另一方面，

$$0\leqslant\mathrm{Var}\left(\frac{X}{\sigma_x}-\frac{Y}{\sigma_y}\right)=\frac{\mathrm{Var}(X)}{\sigma_x^2}+\frac{\mathrm{Var}(Y)}{(-\sigma_y)^2}-2\,\frac{\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y}=2[1-\rho(X,Y)]$$

这意味着

$$\rho(X,Y)\leqslant 1$$

这就完成了式(3-4)的证明.

事实上，由于 $\mathrm{Var}(Z)=0$ 意味着 Z 以概率 1 为一个常量(这一直观的事实将在第 8 章中严格证明)，由式(3-3)的证明可知 $\rho(X,Y)=1$ 意味着 $Y=a+bX$，其中 $b=\sigma_x/\sigma_y>0$，且 $\rho(X,Y)=-1$ 意味着 $Y=a+bX$，其中 $b=-\sigma_x/\sigma_y<0$. 上面的逆命题也成立(留给读者作为练习)：如果 $Y=a+bX$，则 $\rho(X,Y)$等于 1 或-1，且正、负号取决于 b 的符号.

相关系数是表征两个随机变量 X 和 Y 之间线性关系的重要指标，$\rho(X,Y)$接近于 1 或-1表明变量 X 和 Y 的线性相关程度很高，而 $\rho(X,Y)$接近于 0 表示二者的线性相关程度很差. 当$\rho(X,Y)$为正值时，表示变量 Y 随着 X 的增大而增大；当 $\rho(X,Y)$为负值时，表示变量 Y 随着 X 的增大而减小. 如果 $\rho(X,Y)=0$，则称 X 与 Y 不相关(uncorrelated).

例 3f 令 I_A 和 I_B 分别为事件 A 和 B 的示性变量，即

$$I_A=\begin{cases}1 & \text{如果事件 } A \text{ 发生}\\ 0 & \text{否则}\end{cases}$$

$$I_B=\begin{cases}1 & \text{如果事件 } B \text{ 发生}\\ 0 & \text{否则}\end{cases}$$

则

$$E[I_A]=P(A) \quad E[I_B]=P(B) \quad E[I_AI_B]=P(AB)$$

所以

$$\mathrm{Cov}(I_A,I_B)=P(AB)-P(A)P(B)=P(B)[P(A|B)-P(A)]$$

由此我们得到一个相当直观的结果，即事件 A 和 B 的示性变量为正相关、不相关或负相关取

决于 $P(A|B)$大于、等于或小于 $P(A)$. ■ 338

下面的例子将表明样本均值与各变量与样本均值的偏差是不相关的.

例 3g 设 X_1, …, X_n 为独立同分布的随机变量，方差为 σ^2，证明

$$\text{Cov}(X_i-\overline{X},\overline{X})=0$$

解

$$\text{Cov}(X_i-\overline{X},\overline{X}) = \text{Cov}(X_i,\overline{X}) - \text{Cov}(\overline{X},\overline{X}) = \text{Cov}\left(X_i,\frac{1}{n}\sum_{j=1}^{n}X_j\right)-\text{Var}(\overline{X})$$

$$= \frac{1}{n}\sum_{j=1}^{n}\text{Cov}(X_i,X_j)-\frac{\sigma^2}{n}=\frac{\sigma^2}{n}-\frac{\sigma^2}{n}=0$$

倒数第二个等式用到了例 3a 的结果，最后一个等式用到

$$\text{Cov}(X_i,X_j)=\begin{cases}0 & \text{如果 } j\neq i \quad \text{由独立性}\\ \sigma^2 & \text{如果 } j=i \quad \text{由于 } \text{Var}(X_i)=\sigma^2\end{cases}$$

尽管 $\overline{X}$ 和偏差 $X_i-\overline{X}$ 是不相关的，但显然它们是不独立的. 然而在 X_i 为正态随机变量这种特殊情形下，$\overline{X}$ 不仅和单个偏差不相关，而且和偏差序列 $X_j-\overline{X}(j=1, \cdots, n)$都是不相关的，这个结果我们会在第 7.9 节进行论证，同样可以得到，在此种情况下，样本均值 $\overline{X}$ 和样本方差 S^2 也是独立的，且$(n-1)S^2/\sigma^2$ 服从自由度为 $n-1$ 的卡方分布.（有关 S^2 的定义见例 3a.） ■

例 3h 进行 m 次独立试验，每次试验会出现 r 种结果中的任一种的概率分别为 P_1, …, P_r，且$\sum_{i=1}^{r}P_i=1$. 如果用 $N_i(i=1, \cdots, r)$表示 m 次试验中出现第 i 种结果的次数，则 N_1, …, N_r 服从多项分布

$$P\{N_1=n_1, \cdots, N_r=n_r\}=\frac{m!}{n_1!\ n_2!\ \cdots n_r!}P_1^{n_1}P_2^{n_2}\cdots P_r^{n_r}$$

这里$\sum_{i=1}^{r}n_i=m$. 对于 $i\neq j$，可知如果 N_i 很大则 N_j 会变得很小，因此从直觉上认为它们是负相关的. 下面利用命题 3.2(iv)来计算它们的协方差，记 339

$$N_i=\sum_{k=1}^{m}I_i(k) \qquad N_j=\sum_{k=1}^{m}I_j(k)$$

这里

$$I_i(k)=\begin{cases}1 & \text{如果第 } k \text{ 次试验出现第 } i \text{ 个结果}\\ 0 & \text{否则}\end{cases}$$

$$I_j(k)=\begin{cases}1 & \text{如果第 } k \text{ 次试验出现第 } j \text{ 个结果}\\ 0 & \text{否则}\end{cases}$$

由命题 3.2(iv)可知

$$\text{Cov}(N_i,N_j)=\sum_{\ell=1}^{m}\sum_{k=1}^{m}\text{Cov}(I_i(k),I_j(\ell))$$

又当 $k\neq\ell$时，

$$\text{Cov}(I_i(k),I_j(\ell))=0$$

这是由于第 k 次试验的结果和第ℓ次试验的结果是独立的. 而另一方面，

$$\text{Cov}(I_i(\ell),I_j(\ell))=E[I_i(\ell)I_j(\ell)]-E[I_i(\ell)]E[I_j(\ell)]=0-P_iP_j=-P_iP_j$$

上面的等式用到了 $I_i(\ell)I_j(\ell)=0$，因为第ℓ次试验不可能出现两个结果 i 和 j. 故得

$$\mathrm{Cov}(N_i, N_j) = -mP_iP_j$$

这也验证了我们的直觉，即 N_i 和 N_j 是负相关的. ■

7.4 条件数学期望

7.4.1 定义

回想如果随机变量 X 和 Y 为联合离散随机变量，在给定条件 $Y=y$ 且 $P\{Y=y\}>0$ 时，X 的条件概率质量函数定义为

$$p_{X|Y}(x|y)=P\{X=x|Y=y\}=\frac{p(x,y)}{p_Y(y)}$$

因此，在这种情况下，对一切使 $p_Y(y)>0$ 的 y，自然地定义在 $Y=y$ 条件下 X 的条件数学期望为

340

$$E[X \mid Y=y]=\sum_x xP\{X=x \mid Y=y\}=\sum_x xp_{X|Y}(x \mid y)$$

例 4a 如果随机变量 X 和 Y 均服从参数为 n 和 p 的独立二项分布，计算在 $X+Y=m$ 条件下 X 的条件数学期望.

解 首先计算在 $X+Y=m$ 条件下的条件概率质量函数. 对于 $k\leqslant\min(n, m)$，

$$P\{X=k|X+Y=m\}=\frac{P\{X=k,X+Y=m\}}{P\{X+Y=m\}}=\frac{P\{X=k,Y=m-k\}}{P\{X+Y=m\}}=\frac{P\{X=k\}P\{Y=m-k\}}{P\{X+Y=m\}}$$

$$=\frac{\binom{n}{k}p^k(1-p)^{n-k}\binom{n}{m-k}p^{m-k}(1-p)^{n-m+k}}{\binom{2n}{m}p^m(1-p)^{2n-m}}=\frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}$$

这里用到(见第 6 章例 3f) $X+Y$ 服从参数为 $2n$ 和 p 的二项分布，因此在 $X+Y=m$ 条件下 X 的条件分布为超几何分布；于是由例 2g 得到

$$E[X|X+Y=m]=\frac{m}{2}$$ ■

类似地，如果随机变量 X 和 Y 是连续型随机变量，且二者的联合密度函数为 $f(x,y)$，那么在给定条件 $Y=y$ 且对任意 y 有 $f_Y(y)>0$ 时，X 的条件概率密度函数为

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

很自然地，在 $Y=y$ 且 $f_Y(y)>0$ 条件下，X 的条件数学期望为

341

$$E[X \mid Y=y]=\int_{-\infty}^{+\infty} xf_{X|Y}(x \mid y)\mathrm{d}x$$

例 4b 假设随机变量 X 和 Y 的联合密度函数为

$$f(x,y)=\frac{\mathrm{e}^{-x/y}\mathrm{e}^{-y}}{y} \qquad 0<x<\infty, 0<y<\infty$$

计算 $E[X|Y=y]$.

解 首先计算条件密度函数：

$$f_{X|Y}(x \mid y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x}=\frac{(1/y)\mathrm{e}^{-x/y}\mathrm{e}^{-y}}{\int_0^{\infty}(1/y)\mathrm{e}^{-x/y}\mathrm{e}^{-y}\mathrm{d}x}$$

$$= \frac{(1/y)e^{-x/y}}{\int_0^{\infty}(1/y)e^{-x/y}dx} = \frac{1}{y}e^{-x/y}$$

这里在 $Y=y$ 条件下，X 的条件分布恰为均值为 y 的指数分布．故

$$E[X \mid Y=y] = \int_0^{\infty}\frac{x}{y}e^{-x/y}dx = y$$ ■

注 正如条件概率满足和普通概率同样的性质一样，条件数学期望也满足和普通数学期望同样的性质．例如，下面公式

$$E[g(X) \mid Y=y] = \begin{cases} \sum_x g(x)p_{X|Y}(x \mid y) & \text{在离散情况下} \\ \int_{-\infty}^{\infty} g(x)f_{X|Y}(x \mid y)dx & \text{在连续情况下} \end{cases}$$

$$E\left[\sum_{i=1}^{n} X_i \mid Y=y\right] = \sum_{i=1}^{n} E[X_i \mid Y=y]$$

也成立．事实上，已知 $Y=y$ 的条件数学期望可以视为在缩小了的样本空间上的普通数学期望，此缩小了的样本空间仅由所有产生 $Y=y$ 的试验结果组成．

342

7.4.2 计算条件数学期望

以 $E[X|Y]$ 表示随机变量 Y 的函数，其在 $Y=y$ 时的取值为 $E[X|Y=y]$．注意到 $E[X|Y]$ 本身也是一个随机变量，有关条件数学期望的极其重要的性质是：

命题 4.1

$$E[X]=E[E[X|Y]] \tag{4-1}$$

如果 Y 是离散型随机变量，则式(4-1)可表示为

$$E[X] = \sum_y E[X \mid Y=y]P\{Y=y\} \tag{4-1a}$$

如果 Y 是连续型随机变量且具有密度函数 $f_Y(y)$，则式(4-1)可表示为

$$E[X] = \int_{-\infty}^{+\infty} E[X \mid Y=y]f_Y(y)dy \tag{4-1b}$$

下面给出当 X 和 Y 均是离散型随机变量时式(4-1)的证明．

证明 我们必须证明

$$E[X] = \sum_y E[X \mid Y=y]P\{Y=y\} \tag{4-2}$$

式(4-2)的右边可写为

$$\begin{aligned}\sum_y E[X \mid Y=y]P\{Y=y\} &= \sum_y \sum_x xP\{X=x \mid Y=y\}P\{Y=y\} \\ &= \sum_y \sum_x x\frac{P\{X=x,Y=y\}}{P\{Y=y\}}P\{Y=y\} \\ &= \sum_y \sum_x xP\{X=x,Y=y\} \\ &= \sum_x x\sum_y P\{X=x,Y=y\} \\ &= \sum_x xP\{X=x\} = E[X]\end{aligned}$$

■ 343

证毕．

可以这样理解式(4-2)：为计算 $E[X]$，我们采用适当的比例来分配 X 的条件数学期望，使得对每个 $Y=y$，$E[X|Y=y]$的权重值取为其成立的概率值．这是一个很有用的办法，在求随机变量的数学期望时，首先得到其条件数学期望，然后通过条件数学期望来求数学期望．下面的例子说明这种方法的应用．

例 4c 一矿工被困在一个有三个门的矿井里，第一个门通到一个坑道，沿此坑道需 3 小时才能脱险．第二个门通到一个坑道，沿此坑道走 5 小时又回到原被困的矿井．第三个门通向一个坑道，沿此坑道走 7 小时后也回到原被困的矿井．如果矿工等可能地选取三个门中的一个，问平均需要多少小时才能脱险？

解 用 X 来表示矿工脱险所需要的时间(小时)，用 Y 表示他最初选取的门的序号，则

$$\begin{aligned}E[X]&=E[X|Y=1]P\{Y=1\}+E[X|Y=2]P\{Y=2\}+E[X|Y=3]P\{Y=3\}\\&=\frac{1}{3}(E[X|Y=1]+E[X|Y=2]+E[X|Y=3])\end{aligned}$$

然而，

$$\begin{aligned}E[X|Y=1]&=3\\E[X|Y=2]&=5+E[X]\\E[X|Y=3]&=7+E[X]\end{aligned}\tag{4-3}$$

下面解释式(4-3)成立的理由，对于 $E[X|Y=2]$的理由如下：如果矿工选择第二个门脱险，则经过 5 小时又回到原出发点，故选择这条途径的平均脱险时间为 $E[X|Y=2]=5+E[X]$，第三个门的情况类似．因此

$$E[X]=\frac{1}{3}(3+5+E[X]+7+E[X])$$

得到

$$E[X]=15$$ ■

例 4d(随机变量的随机数之和的数学期望) 假设一天内进入一家商店的顾客人数为均值为 50 的随机变量，进一步假设每个顾客消费的钱数是均值为 8 美元的独立随机变量，且每个顾客消费的钱数与一天内进入商店的顾客数也是独立的，求某天顾客总消费钱数的期望值．

344

解 令 N 表示进入这家商店的顾客人数，令 X_i 表示第 i 个顾客在这家商店消费的钱数，则所有顾客消费的钱数为 $\sum_{i=1}^{N}X_i$．现在，有

$$E\left[\sum_{i=1}^{N}X_i\right]=E\left[E\left[\sum_{i=1}^{N}X_i \mid N\right]\right]$$

而

$$\begin{aligned}E\left[\sum_{i=1}^{N}X_i \mid N=n\right]&=E\left[\sum_{i=1}^{N}X_i \mid N=n\right]\\&=E\left[\sum_{i=1}^{N}X_i\right] \qquad \text{由 } X_i \text{ 与 } N \text{ 的独立性}\\&=nE[X] \qquad \text{其中 } E[X]=E[X_i]\end{aligned}$$

这意味着

$$E\left[\sum_{i=1}^{N}X_i \mid N\right]=NE[X]$$

因此

$$E\Big[\sum_{i=1}^{N}X_i\Big]=E[NE[X]]=E[N]E[X]$$

故由上面的结果可知，某天顾客在这家商店的总消费钱数的期望值为 $50\times8=400$ 美元. ■

例 4e 掷骰子游戏的规则为：首先掷一对骰子，如果显示的点数之和为 2，3 或 12，则游戏者输一局；如果点数之和为 7 或 11，则游戏者赢一局；如果点数之和为其他数字 i，则游戏者继续掷，直到点数之和为 7 或 i，如果是 7，则游戏者输一局，如果是 i，则游戏者赢一局. 用 R 表示一局游戏中分出输赢所需要的掷骰子次数. 求

(a) $E[R]$.

(b) $E[R|$游戏者赢$]$.

(c) $E[R|$游戏者输$]$.

解 用 W 表示游戏者赢，L 表示游戏者输. 如果用 P_i 表示点数之和为 i 的概率，则

$$P_i=P_{14-i}=\frac{i-1}{36},i=2,\cdots,7$$

为计算 $E[R]$，假设首次掷的点数之和为 S，则有

$$E[R]=\sum_{i=2}^{12}E[R\mid S=i]P_i$$

345

而

$$E[R|S=i]=\begin{cases}1 & \text{如果 } i=2,3,7,11,12\\ 1+\dfrac{1}{P_i+P_7} & \text{其他}\end{cases}$$

如果首次显示点数之和为 i，游戏继续进行直到显示点数之和为 i 或 7，则直到游戏结束所需要掷的次数服从参数为 P_i+P_7 的几何分布. 因此

$$E[R]=1+\sum_{i=4}^{6}\frac{P_i}{P_i+P_7}+\sum_{i=8}^{10}\frac{P_i}{P_i+P_7}=1+2\Big(\frac{3}{9}+\frac{4}{10}+\frac{5}{11}\Big)=3.376$$

为确定 $E[R/W]$，我们用 p 表示游戏者赢的概率. 由关于 S 的条件概率得

$$p=\sum_{i=2}^{12}P\{W\mid S=i\}P_i=P_7+P_{11}+\sum_{i=4}^{6}\frac{P_i}{P_i+P_7}P_i+\sum_{i=8}^{10}\frac{P_i}{P_i+P_7}P_i=0.493$$

上面等式用到在获得点数之和为 7 之前获得点数之和为 i 的概率为$\frac{P_i}{P_i+P_7}$，下面给出在游戏者赢的情况下 S 的条件概率质量函数. 令 $Q_i=P\{S=i/W\}$，则

$$Q_2=Q_3=Q_{12}=0,Q_7=\frac{P_7}{p},Q_{11}=\frac{P_{11}}{p}$$

对于 $i=4,5,6,8,9,10$，

$$Q_i=\frac{P\{S=i,W\}}{P\{W\}}=\frac{P_iP\{W|S=i\}}{p}=\frac{P_i^2}{p(P_i+P_7)}$$

而

$$E[R\mid W]=\sum_i E[R\mid W,S=i]Q_i$$

然而如第 6 章例 2j 所述，假设首次出现的点数之和为 i，后面需要的游戏次数与游戏结果(不管是赢还是输)是独立的.（很容易看出这一点，首先注意以首次出现的点数之和为 i 作为条件，

346

游戏结果与后面需要的游戏次数是独立的，然后用独立性的对称性质——如果事件 A 与事件 B 是独立的，则事件 B 与事件 A 也是独立的.）因此，我们看到

$$E[R \mid W] = \sum_i E[R \mid S = i]Q_i = 1 + \sum_{i=4}^{6} \frac{Q_i}{P_i + P_7} + \sum_{i=8}^{10} \frac{Q_i}{P_i + P_7} = 2.938$$

尽管可以采用同样的办法来求 $E[R|L]$，然而也可采用下面更简单的方法：

$$E[R]=E[R|W]p+E[R|L](1-p)$$

这意味着

$$E[R|L]=\frac{E[R]-E[R|W]p}{1-p}=3.801$$ ■

例 4f 一个坛子里装有 a 个白球和 b 个黑球，每次随机地抽取一个球直到第一个白球被取到，求取出的黑球数的期望值.

解 这个问题在例 2m 中已经讨论过，下面应用条件期望来解决. 令 X 表示需要取出的黑球数，为了更明确地突出是 a 和 b 的函数，令 $M_{a,b}=E[X]$，给出其关于第一次取出的类型的条件期望表达式，即定义

$$Y=\begin{cases}1 & \text{如果第一次取出的是白球}\\0 & \text{如果第一次取出的是黑球}\end{cases}$$

得

$$M_{a,b}=E[X]=E[X|Y=1]P\{Y=1\}+E[X|Y=0]P\{Y=0\}$$

而

$$E[X|Y=1]=0 \tag{4-4}$$

$$E[X|Y=0]=1+M_{a,b-1} \tag{4-5}$$

为了理解式(4-4)和式(4-5)的含义，假设第一次取出的球是黑球，则以后的情况和原来坛子里有 a 个白球与 $b-1$ 个黑球是一样的，这就是式(4-5)的含义.

347

由于 $P\{Y=0\}=b/(a+b)$，可知

$$M_{a,b}=\frac{b}{a+b}[1+M_{a,b-1}]$$

由 $M_{a,0}=0$ 得到

$$M_{a,1}=\frac{1}{a+1}[1+M_{a,0}]=\frac{1}{a+1}$$

$$M_{a,2}=\frac{2}{a+2}[1+M_{a,1}]=\frac{2}{a+2}\left[1+\frac{1}{a+1}\right]=\frac{2}{a+1}$$

$$M_{a,3}=\frac{3}{a+3}[1+M_{a,2}]=\frac{3}{a+3}\left[1+\frac{2}{a+1}\right]=\frac{3}{a+1}$$

由归纳法可得

$$M_{a,b}=\frac{b}{a+1}$$ ■

同样可以通过条件期望来求随机变量的方差，举例如下.

例 4g（几何分布的方差） 在一系列独立试验中，每次试验成功的概率为 p，用 N 表示首次成功所需要的试验次数，求 $\mathrm{Var}(N)$.

解 若第一次试验成功我们记 $Y=1$，否则记 $Y=0$. 现在，有

$$\mathrm{Var}(N)=E[N^2]-(E[N])^2$$

为计算 $E[N^2]$，我们取给定 Y 的条件期望：

$$E[N^2]=E[E[N^2|Y]]$$

然而

$$E[N^2|Y=1]=1$$
$$E[N^2|Y=0]=E[(1+N)^2]$$

这两个式子成立是因为，如果第一次试验成功则 $N=1$，因此 $N^2=1$. 另一方面，如果第一次试验失败，那么直到成功所需要的试验次数与 1(第一次试验失败)加上下面试验所需要的次数有相同的分布. 由于后者与 N 有相同的分布，得 $E[N^2|Y=0]=E[(1+N)^2]$. 因此有

$$E[N^2]=E[N^2|Y=1]P\{Y=1\}+E[N^2|Y=0]P\{Y=0\}$$
$$=p+(1-p)E[(1+N)^2]=1+(1-p)E[2N+N^2]$$

348

而由第 4 章例 9b 可知 $E[N]=1/p$，故

$$E[N^2]=1+\frac{2(1-p)}{p}+(1-p)E[N^2]$$

或

$$E[N^2]=\frac{2-p}{p^2}$$

因此，

$$\mathrm{Var}(N)=E[N^2]-(E[N])^2=\frac{2-p}{p^2}-\left(\frac{1}{p}\right)^2=\frac{1-p}{p^2}$$ ■

下面的例子将确定服从(0, 1)上均匀分布的随机变量之和大于 1 所需要的取值个数的数学期望. 令人感到意外的是其结果为 e.

例 4h 设 U_1，U_2，… 表示服从(0, 1) 上均匀分布的独立随机变量序列，且 $N=\min\left\{n: \sum_{i=1}^{n}U_i>1\right\}$，求 $E[N]$.

解 通过一个更一般的结论来解决上面的问题. 对于 $x\in[0, 1]$，令

$$N(x)=\min\left\{n: \sum_{i=1}^{n}U_i>x\right\}$$

并令

$$m(x)=E[N(x)]$$

即 $N(x)$表示加到其和大于 x 的服从(0, 1)上均匀分布的随机变量的个数，$m(x)$为其数学期望，我们将推出一个给定 U_1 时 $m(x)$的条件期望的等式. 由式(4-1b)得

$$m(x)=\int_0^1 E[N(x)\mid U_1=y]\mathrm{d}y \tag{4-6}$$

现在，有

$$E[N(x)|U_1=y]=\begin{cases}1 & \text{如果 } y>x\\ 1+m(x-y) & \text{如果 } y\leqslant x\end{cases} \tag{4-7}$$

对于 $y>x$ 上式显然成立，对于 $y\leqslant x$，由于如果第一个均匀随机变量的取值是 y，则这时剩余的均匀随机变量的个数与最初使取值大于 $x-y$ 所需要的均匀随机变量数相同，所以将式(4-7)代入式(4-6)可得

349

$$m(x)=1+\int_0^x m(x-y)\mathrm{d}y=1+\int_0^x m(u)\mathrm{d}u \qquad \text{令 } u=x-y$$

对上式两边求导可得

$$m'(x)=m(x)$$

或

$$\frac{m'(x)}{m(x)}=1$$

对上式积分得

$$\log[m(x)]=x+c$$

或

$$m(x)=ke^x$$

由 $m(0)=1$ 得 $k=1$，即

$$m(x)=e^x$$

因此需要加到其和大于 1 的服从(0，1)上均匀分布的随机变量的个数的数学期望为 $m(1)=e$. ■

7.4.3 通过设置条件计算概率

不仅可以用适当的随机变量作为最初的条件来计算数学期望，而且还可以用这个方法来计算概率. 为此，令 E 为一任意事件，定义示性变量 X 为

$$X=\begin{cases}1 & \text{如果 } E \text{ 发生}\\ 0 & \text{如果 } E \text{ 不发生}\end{cases}$$

由上面定义可知

$$E[X]=P(E)$$

$$E[X|Y=y]=P(E|Y=y) \qquad \text{对任意随机变量 } Y$$

因此，由式(4-1a)和式(4-1b)得到

$$\begin{aligned}P(E) &= \sum_y P(E \mid Y=y)P(Y=y) && \text{如果 } Y \text{ 是离散型随机变量}\\ &= \int_{-\infty}^{+\infty} P(E \mid Y=y)f_Y(y)\mathrm{d}y && \text{如果 } Y \text{ 是连续型随机变量}\end{aligned} \tag{4-8}$$

350

假设 Y 是一离散型随机变量，取值为 y_1，y_2，…，y_n 中的一个，用 $F_i(i=1, \cdots, n)$来表示事件 $F_i=\{Y=y_i\}$，则式(4-8)可表示为

$$P(E)=\sum_{i=1}^{n}P(E \mid F_i)P(F_i)$$

这里 F_1，…，F_n 是同一样本空间上的互不相容事件且其并是样本空间.

例 4i(最优奖品问题) 假设 n 个不同的奖品顺次出现，当我们看到一个奖品时，必须马上做出决定是接受它还是拒绝它，然后考虑下一个奖品. 我们决定是否接受一个奖品的依据是我们知道的唯一信息，即现在出现的奖品与我们已经看到的奖品比较的相对等级. 例如，当看到第五个奖品时，就可以将这个奖品与前四个奖品进行比较然后做出决定. 假设一个奖品一旦被拒绝后将不再被选择，我们的目标是获得最优奖品的最大概率. 假设奖品的所有 $n!$ 种排列是等可能的，我们应该如何去做?

解 令人惊奇的是我们能做得非常好. 为了看到这一点，固定 k，$0\leqslant k\leqslant n$，考虑的策略是拒绝前 k 个奖品，然后接受第一个比前 k 个奖品优的那个奖品，用 $P_k(B)$表示当采用上述策略时能取到最优奖品的概率，为计算这个概率，对最优奖品所处的次序值 X 取条件概率，即

$$P_k(B)=\sum_{i=1}^{n}P_k(B\mid X=i)P(X=i)=\frac{1}{n}\sum_{i=1}^{n}P_k(B\mid X=i)$$

如果最优奖品包含在前 k 个被拒绝的奖品内，则不会再取到最优奖品，即

$$P_k(B|X=i)=0\qquad 如果\ i\leqslant k$$

另一方面，如果最优奖品在第 $i(i>k)$ 个位置出现，且前 $i-1$ 个奖品中的最优奖品出现在前 k 个奖品内(于是第 $k+1$, $k+2$, …, $i-1$ 个奖品不会被取到)，则我们会得到最优奖品. 当最优奖品出现在第 i 个位置时，很明显其余奖品的排列顺序也是等可能的，这意味着前 $i-1$ 个奖品均等可能地成为这批中的最优奖品，因此

$$\begin{aligned}P_k(B|X=i)&=P\{前\ i-1\ 个奖品中的最优奖品出现在前\ k\ 个中|X=i\}\\&=\frac{k}{i-1}\qquad 如果\ i>k\end{aligned}$$

351

从上面式子得到

$$P_k(B)=\frac{k}{n}\sum_{i=k+1}^{n}\frac{1}{i-1}\approx\frac{k}{n}\int_{k+1}^{n}\frac{1}{x-1}\mathrm{d}x=\frac{k}{n}\log\left(\frac{n-1}{k}\right)\approx\frac{k}{n}\log\left(\frac{n}{k}\right)$$

下面考虑函数

$$g(x)=\frac{x}{n}\log\left(\frac{n}{x}\right)$$

则

$$g'(x)=\frac{1}{n}\log\left(\frac{n}{x}\right)-\frac{1}{n}$$

故

$$g'(x)=0\Rightarrow\log\left(\frac{n}{x}\right)=1\Rightarrow x=\frac{n}{\mathrm{e}}$$

由于 $P_k(B)\approx g(k)$，可知我们可以采用的最优策略为观察前 n/e 个奖品，然后接受接下来的第一个比前 n/e 个优的奖品. 由于 $g(n/\mathrm{e})=1/\mathrm{e}$，所以这种策略能取到最优奖品的概率为 $1/\mathrm{e}=0.367\,88$. ■

注 可能很多读者会对得到最优奖品的概率感到吃惊，一般会认为当 n 很大时，取到最优奖品的概率趋于 0. 然而，即使不用仔细计算，也不会想到结果会是这个值. 考虑观察前一半的奖品，然后取接下来的第一个比前一半奖品优的那个奖品. 实际上选中一个奖品的概率等于最优奖品在后一半中的概率，即$\frac{1}{2}$，另外，假设一个奖品被选中，则这个奖品比已经出现的至少$\frac{n}{2}$个奖品优，至少有$\frac{1}{2}$的可能性成为全部奖品中的最优奖品. 因此，采用观察前一半奖品的策略至少有$\frac{1}{4}$的可能性取到最优奖品.

例 4j 设 U 为服从(0, 1)上均匀分布的随机变量，在 $U=p$ 的条件下，X 的条件分布是参数为 n 和 p 的二项分布，求 X 的概率质量函数.

352

解 对 U 的取值求条件概率，有

$$\begin{aligned}P\{X=i\}&=\int_0^1 P\{X=i\mid U=p\}f_U(p)\mathrm{d}p\\&=\int_0^1 P\{X=i\mid U=p\}\mathrm{d}p=\frac{n!}{i!(n-i)!}\int_0^1 p^i(1-p)^{n-i}\mathrm{d}p\end{aligned}$$

由等式

$$\int_0^1 p^i(1-p)^{n-i}\mathrm{d}p = \frac{i!(n-i)!}{(n+1)!}$$

成立(证明过程见 6.6 节)得到

$$P\{X=i\}=\frac{1}{n+1} \qquad i=0,1,\cdots,n$$

这里也得到一个令人惊奇的结果：抛一枚硬币，正面朝上的概率服从(0, 1)上的均匀分布，抛 n 次，则正面朝上的次数等可能地是 0, 1, …, n 中的任何一个.

由于上面的条件分布具有很好的形式，所以可以从另一个角度来考虑为什么会出现上面的结果. 令 $U, U_1, \cdots, U_n$ 为 $n+1$ 个相互独立的服从(0, 1)上均匀分布的随机变量，用 X 来表示 $U_1, \cdots, U_n$ 中取值比 U 的取值小的个数. 由于随机变量 $U, U_1, \cdots, U_n$ 的分布相同，故 U 等可能地是最小值、次小值或最大值，即 X 等可能地取值为 0, 1, …, n 中的一个. 而在给定 $U=p$ 的情况下，U_i 比 U 小的个数服从参数为 n 和 p 的二项分布，于是得到上面的结果. ■

例 4k 设 X 和 Y 为独立的连续型随机变量，其密度函数分别为 f_X 和 f_Y，计算 $P\{X<Y\}$.

解 以 Y 的取值为条件可得

$$\begin{aligned}
P\{X<Y\} &= \int_{-\infty}^{+\infty} P\{X<Y \mid Y=y\} f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} P\{X<y \mid Y=y\} f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} P\{X<y\} f_Y(y)\mathrm{d}y \qquad \text{由独立性} \\
&= \int_{-\infty}^{+\infty} F_X(y) f_Y(y)\mathrm{d}y
\end{aligned}$$

353

其中

$$F_X(y) = \int_{-\infty}^{y} f_X(x)\mathrm{d}x$$

■

例 4l 设 X 和 Y 为独立的连续型随机变量，求 $X+Y$ 的分布.

解 以 Y 的取值为条件可得

$$\begin{aligned}
P\{X+Y<a\} &= \int_{-\infty}^{+\infty} P\{X+Y<a \mid Y=y\} f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} P\{X+y<a \mid Y=y\} f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} P\{X<a-y\} f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} F_X(a-y) f_Y(y)\mathrm{d}y
\end{aligned}$$

这里 $F_X(y)$ 的定义同上. ■

7.4.4 条件方差

正如前面在给定 Y 值的条件下定义 X 的条件期望一样，也可以在给定 $Y=y$ 的条件下定义 X 的方差，其定义如下：

$$\mathrm{Var}(X|Y) \equiv E[(X-E[X|Y])^2 | Y]$$

即 $\mathrm{Var}(X|Y)$ 等于在给定 Y 值的情况下，X 与条件期望之间差的平方的期望，换句话说，

$\mathrm{Var}(X|Y)$的定义方式和通常的方差定义是类似的，所不同的是所有的期望是在 Y 取值已知的条件下.

方差 $\mathrm{Var}(X)$(随机变量 X 的无条件方差)与条件方差 $\mathrm{Var}(X|Y)$之间有很有用的关系．在 Y 给定条件下的条件方差常用来求方差 $\mathrm{Var}(X)$. 为得到这一关系，首先注意通过与 $\mathrm{Var}(X)=E[X^2]-(E[X])^2$ 同样的推理有

$$\mathrm{Var}(X|Y)=E[X^2|Y]-(E[X|Y])^2$$

故

$$E[\mathrm{Var}(X|Y)]=E[E[X^2|Y]]-E[(E[X|Y])^2]=E[X^2]-E[(E[X|Y])^2] \tag{4-9}$$

此外，由 $E[E[X|Y]]=E[X]$得到

$$\mathrm{Var}(E[X|Y])=E[(E[X|Y])^2]-(E[X])^2 \tag{4-10}$$

因此，式(4-9)和式(4-10)两边相加，可得到下面的命题.

354

命题 4.2(条件方差公式)

$$\mathrm{Var}(X)=E[\mathrm{Var}(X|Y)]+\mathrm{Var}(E[X|Y])$$

例 4m 假设任意时刻 t 到达火车站的人数服从均值为 λt 的泊松分布，且火车到站的时刻(与乘客到站的时刻独立)服从$(0, T)$上的均匀分布，求火车到站时上车的人数的期望和方差.

解 对于任意时刻 $t(t\geqslant 0)$，用 $N(t)$表示到 t 时刻为止到站的人数，用 Y 表示火车到达的时刻，我们关心的是随机变量 $N(Y)$，以 Y 的取值为条件得到

$$\begin{aligned} E[N(Y)|Y=t] &= E[N(t)|Y=t] \\ &= E[N(t)] \qquad \text{由 } Y \text{ 与 } N(t) \text{的独立性} \\ &= \lambda t \qquad \text{由于 } N(t) \text{服从均值为 } \lambda t \text{ 的泊松分布} \end{aligned}$$

因此

$$E[N(Y)|Y]=\lambda Y$$

取期望可得

$$E[N(Y)]=\lambda E[Y]=\frac{\lambda T}{2}$$

为获得 $\mathrm{Var}(N(Y))$，用条件方差公式：

$$\begin{aligned} \mathrm{Var}(N(Y)|Y=t) &= \mathrm{Var}(N(t)|Y=t) \\ &= \mathrm{Var}(N(t)) \qquad \text{由独立性} \\ &= \lambda t \end{aligned}$$

故

$$\mathrm{Var}(N(Y)|Y)=\lambda Y$$
$$E[N(Y)|Y]=\lambda Y$$

因此，由条件期望公式得

$$\mathrm{Var}(N(Y))=E[\lambda Y]+\mathrm{Var}(\lambda Y)=\lambda\frac{T}{2}+\lambda^2\frac{T^2}{12}$$

上式用到 $\mathrm{Var}(T)=\frac{T^2}{12}$. ■

355

例 4n(随机变量的随机数之和的方差) 设 X_1，X_2，… 为一独立同分布的随机变量序列，N 是只取非负整数值的随机变量，且与 $X_i(i\geqslant 1)$ 相互独立，试计算 $\mathrm{Var}\left(\sum_{i=1}^{N}X_i\right)$.

解 取 N 为条件，有

$$E\left[\sum_{i=1}^{N} X_i \middle| N\right] = NE[X]$$

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i \middle| N\right) = N\mathrm{Var}(X)$$

由于对给定的 N，$\sum_{i=1}^{N} X_i$ 是数目固定的独立随机变量之和，因此它的数学期望和方差正好是它们各自的数学期望和方差的和．因此，应用条件方差公式得

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = E[N]\mathrm{Var}(X) + (E[X])^2\mathrm{Var}(N)$$ ■

7.5　条件数学期望与预测

我们有时会遇到如下的情况：已经得到随机变量 X 的观察值，根据这些观察值试图预测另一个随机变量 Y 的取值．令 $g(X)$ 来表示这个预测，即若 X 的观察值等于 x，那么 $g(x)$ 就是对随机变量 Y 的预测值．明确地说，就是选择一个函数 g，使得 $g(X)$ 尽可能地接近 Y．选择函数 g 的一种准则是使得 $E[(Y-g(X))^2]$ 的取值达到最小，我们将证明在这种准则下，对 Y 的最优可能预测为 $g(X)=E[Y|X]$．

命题 5.1

$$E[(Y-g(X))^2] \geqslant E[(Y-E[Y|X])^2]$$

证明

$$\begin{aligned} E[(Y-g(X))^2|X] &= E[(Y-E[Y|X]+E[Y|X]-g(X))^2|X] \\ &= E[(Y-E[Y|X])^2|X]+E[(E[Y|X]-g(X))^2|X] \\ &\quad +2E[(Y-E[Y|X])(E[Y|X]-g(X))|X] \end{aligned} \tag{5-1}$$

[356]

然而在给定 X 后，$E[Y|X]-g(X)$ 作为 X 的函数，可以看作一个常数，因此

$$\begin{aligned} E[(Y-E[Y|X])(E[Y|X]-g(X))|X] &= (E[Y|X]-g(X))E[Y-E[Y|X]|X] \\ &= (E[Y|X]-g(X))(E[Y|X]-E[Y|X])=0 \end{aligned} \tag{5-2}$$

由式(5-1)和式(5-2)可得

$$E[(Y-g(X))^2|X] \geqslant E[(Y-E[Y|X])^2|X]$$

对上式两边取期望可得命题成立． ■

注　下面给出命题 5.1 的另一个尽管不太严格但很直观的讨论．很显然 $E[(Y-c)^2]$ 在 $c=E[Y]$ 处取得最小值(见理论练习 1)，因此在没有任何数据可以利用的情况下，按照最小化均方误差原则，对 Y 最好的预测是其均值．另一方面，如果随机变量 X 的观察值为 x，则此时的最优预测同没有数据时采用的方法一样，只是所有的概率和数学期望都要在 $X=x$ 条件下来考虑，因此此时的最优预测是 Y 在 $X=x$ 条件下的条件期望，即命题 5.1 的结论．

例 5a　假设身高为 x 英寸[1]的父亲，其儿子的身高(英寸)服从均值为 $x+1$、方差为 4 的正态分布．假设一父亲身高 6 英尺[2]，则对其儿子身高的最佳预测值为多少？

[1] 英寸的单位符号为 in，1 in=0.025 4 m．——编辑注

[2] 英尺的单位符号为 ft，1 ft=0.304 8 m．——编辑注

解 通常可以建立以下模型：

$$Y=X+1+e$$

其中，e 与 X 独立，且服从均值和方差分别为 0 和 4 的标准正态分布．X 和 Y 分别表示父亲和儿子的身高，则最优的预测值 $E[Y|X=72]$ 为

$$\begin{aligned}E[Y|X=72]&=E[X+1+e|X=72]=73+E[e|X=72]\\&=73+E[e] \qquad \text{由独立性}\\&=73\end{aligned}$$

■

例 5b 假设一个值为 s 的信号从位置 A 发出，则在位置 B 接收到的信号值服从参数为 $(s, 1)$ 的正态分布．设从位置 A 发出的信号 S 服从参数为 (μ,σ^2) 的正态分布，假设在位置 B 接收到的信号值 $R=r$，则对位置 A 发出的信号值的最优预测为多少？

[357]

解 首先计算给定 R 的条件下 S 的条件密度函数：

$$f_{S|R}(s|r)=\frac{f_{S,R}(s,r)}{f_R(r)}=\frac{f_S(s)f_{R|S}(r|s)}{f_R(r)}=K\mathrm{e}^{-(s-\mu)^2/2\sigma^2}\mathrm{e}^{-(r-s)^2/2}$$

这里 K 不依赖于 s. 现在，有

$$\begin{aligned}\frac{(s-\mu)^2}{2\sigma^2}+\frac{(r-s)^2}{2}&=s^2\left(\frac{1}{2\sigma^2}+\frac{1}{2}\right)-\left(\frac{\mu}{\sigma^2}+r\right)s+C_1=\frac{1+\sigma^2}{2\sigma^2}\left[s^2-2\left(\frac{\mu+r\sigma^2}{1+\sigma^2}\right)s\right]+C_1\\&=\frac{1+\sigma^2}{2\sigma^2}\left(s-\frac{(\mu+r\sigma^2)}{1+\sigma^2}\right)^2+C_2\end{aligned}$$

这里 C_1，C_2 不依赖于 s. 因此

$$f_{S|R}(s|r)=C\exp\left\{\frac{-\left[s-\dfrac{(\mu+r\sigma^2)}{1+\sigma^2}\right]^2}{2\left(\dfrac{\sigma^2}{1+\sigma^2}\right)}\right\}$$

这里 C 不依赖于 s. 于是，我们可以断定在接到的信号为 r 的情况下发出信号 S 的条件分布是正态分布，均值和方差分别为

$$E[S|R=r]=\frac{\mu+r\sigma^2}{1+\sigma^2}$$

$$\mathrm{Var}(S|R=r)=\frac{\sigma^2}{1+\sigma^2}$$

因此，由命题 5.1 可知在接收信号值 r 一定的情况下，按照最小化均方误差原则，对发出信号的最优预测为

$$E[S|R=r]=\frac{1}{1+\sigma^2}\mu+\frac{\sigma^2}{1+\sigma^2}r$$

将条件期望写成上述表达式是有益的，因为它表明条件期望等于信号的先验均值 μ 与接收信号 r 的加权平均，其各自的权重之比为 1（当发送信号为 s 时收到信号的条件方差）比 σ^2（发送信号的方差）. ■

例 5c 在数字信号处理过程中，原始的连续数据 X 经过处理后将被量化为一序列离散化的数字，其基本原理为：首先选取一递增的数列 $a_i(i=0,\pm1,\pm2,\cdots)$ 使得 $\lim\limits_{i\to+\infty}a_i=+\infty$，$\lim\limits_{i\to-\infty}a_i=-\infty$，如果 X 界于某个区间 $(a_i,a_{i+1}]$ 中，就得到处理后的数值 y_i. 我们用 Y 表示处理后的数值．即 $Y=y_i$，如果 $a_i<X\leqslant a_{i+1}$，则 Y 的概率分布为

[358]

$$P\{Y=y_i\}=F_X(a_{i+1})-F_X(a_i)$$

下面考虑如何选取 $y_i(i=0,\pm1,\pm2,\cdots)$ 使得原始数据与处理后的数据之间的均方误差 $E[(X-Y)^2]$ 最小.

(a) 求最优的 $y_i(i=0,\pm1,\pm2,\cdots)$ 使得原始数据与处理后的数据之间的均方误差 $E[(X-Y)^2]$ 最小.

(b) 对于最优的量化数值 Y 证明 $E[X]=E[Y]$，即处理前后的数据具有相同的均值.

(c) 对于最优的量化数值 Y 证明 $\mathrm{Var}(Y)=\mathrm{Var}(X)-E[(X-Y)^2]$.

解　(a)对于任意的量化值 Y，基于 Y 的取值可得

$$E[(X-Y)^2]=\sum_i E[(X-y_i)^2 \mid a_i<X\leqslant a_{i+1}]P\{a_i<X\leqslant a_{i+1}\}$$

如果 $a_i<X\leqslant a_{i+1}$ 时记 $I=i$，则

$$E[(X-y_i)^2 \mid a_i<X\leqslant a_{i+1}]=E[(X-y_i)^2 \mid I=i]$$

由命题 5.1，当

$$y_i=E[X \mid I=i]=E[X \mid a_i<X\leqslant a_{i+1}]=\int_{a_i}^{a_{i+1}}\frac{xf_X(x)\mathrm{d}x}{F_X(a_{i+1})-F_X(a_i)}$$

时上式取最小值. 由于最优的量化值为 $Y=E[X \mid I]$，因此

(b) $E[Y]=E[X]$

(c)
$$\begin{aligned}\mathrm{Var}(X)&=E[\mathrm{Var}(X \mid I)]+\mathrm{Var}(E[X \mid I])=E[E[(X-Y)^2 \mid I]]+\mathrm{Var}(Y)\\&=E[(X-Y)^2]+\mathrm{Var}(Y)\end{aligned}$$

■

有时，随机变量 X 和 Y 的联合概率分布函数是不完全知道的，或者即使知道，条件期望 $E[Y \mid X=x]$ 也是难以处理的. 然而，如果我们知道 X 和 Y 的均值和方差与 X 和 Y 的相关系数，就至少可以确定 Y 关于 X 的最佳线性预测.

359

为获得 Y 关于 X 的最佳线性预测，我们需要选择常数 a 与 b，使得 $E[(Y-(a+bX))^2]$ 取得最小值. 现在，有

$$\begin{aligned}E[(Y-(a+bX))^2]&=E[Y^2-2aY-2bXY+a^2+2abX+b^2X^2]\\&=E[Y^2]-2aE[Y]-2bE[XY]+a^2+2abE[X]+b^2E[X^2]\end{aligned}$$

对上式求偏导得

$$\begin{aligned}&\frac{\partial}{\partial a}E[(Y-a-bX)^2]=-2E[Y]+2a+2bE[X]\\&\frac{\partial}{\partial b}E[(Y-a-bX)^2]=-2E[XY]+2aE[X]+2bE[X^2]\end{aligned}\tag{5-3}$$

令上式分别等于 0，解关于 a 和 b 的方程可得

$$\begin{aligned}b&=\frac{E[XY]-E[X]E[Y]}{E[X^2]-(E[X])^2}=\frac{\mathrm{Cov}(X,Y)}{\sigma_x^2}=\rho\frac{\sigma_y}{\sigma_x}\\a&=E[Y]-bE[X]=E[Y]-\frac{\rho\sigma_y E[X]}{\sigma_x}\end{aligned}\tag{5-4}$$

这里 ρ 为 X 与 Y 的相关系数，$\sigma_x^2=\mathrm{Var}(X)$，$\sigma_y^2=\mathrm{Var}(Y)$，则很显然式(5-4)中的取值能使 $E[(Y-a-bX)^2]$ 达到最小，因此 Y 关于 X 的最佳线性预测为

$$\mu_y+\frac{\rho\sigma_y}{\sigma_x}(X-\mu_x)$$

其中 $\mu_x=E[X]$，$\mu_y=E[Y]$.

这个预测的均方误差为

$$E\left[\left(Y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(X-\mu_x)\right)^2\right]$$

$$=E[(Y-\mu_y)^2]+\rho^2\frac{\sigma_y^2}{\sigma_x^2}E[(X-\mu_x)^2]-2\rho\frac{\sigma_y}{\sigma_x}E[(Y-\mu_y)(X-\mu_x)]$$

$$=\sigma_y^2+\rho^2\sigma_y^2-2\rho^2\sigma_y^2=\sigma_y^2(1-\rho^2) \tag{5-5}$$

由式(5-5)可知当 ρ 接近于 $+1$ 或 -1 时，最佳线性预测的均方误差接近于 0.

例 5d 在给定 X 条件下 Y 的条件期望与 X 成线性关系，从而 Y 关于 X 的最佳线性预测在所有预测中为最优的例子是，X 和 Y 具有二元正态分布，这时其联合密度函数为 [360]

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

在 $X=x$ 条件下 Y 的条件密度函数为

$$f_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(y-\mu_y-\frac{\rho\sigma_y}{\sigma_x}(x-\mu_x)\right)^2\right\}$$

这留给读者去证明. 因此在 $X=x$ 条件下，Y 的条件分布函数是均值为

$$E[Y|X=x]=\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$$

方差为 $\sigma_y^2(1-\rho^2)$ 的正态分布函数. ■

7.6 矩母函数

随机变量 X 的矩母函数 $M(t)$ 定义为：对一切实数 t，

$$M(t)=E[e^{tX}]=\begin{cases}\sum_x e^{tx}p(x) & \text{如果 } X \text{ 是离散型且有概率质量函数 } p(x)\\ \int_{-\infty}^{\infty} e^{tx}f(x)dx & \text{如果 } X \text{ 是连续型且有密度函数 } f(x)\end{cases}$$

我们称 $M(t)$ 为矩母函数，是因为随机变量的各阶矩都可以通过对 $M(t)$ 逐次求导，并取其在 $t=0$ 处的导数值而获得. 例如，

$$M'(t)=\frac{d}{dt}E[e^{tX}]=E\left[\frac{d}{dt}(e^{tX})\right]=E[Xe^{tX}] \tag{6-1}$$

在上式中我们假定了求导数与求数学期望这两种运算可以交换，也就是说，假定在离散型情形有 [361]

$$\frac{d}{dt}\left[\sum_x e^{tx}p(x)\right]=\sum_x\frac{d}{dt}[e^{tx}p(x)]$$

在连续型情形有

$$\frac{d}{dt}\left[\int e^{tx}f(x)dx\right]=\int\frac{d}{dt}[e^{tx}f(x)]dx$$

这个假定差不多总能满足. 事实上，本书中的所有分布函数都适合这个假定. 因此，在式(6-1)中令 $t=0$，即得

$$M'(0)=E[X]$$

同样，

$$M''(t)=\frac{d}{dt}M'(t)=\frac{d}{dt}E[Xe^{tX}]=E\left[\frac{d}{dt}(Xe^{tX})\right]=E[X^2e^{tX}]$$

因此可得

$$M''(0)=E[X^2]$$

一般地，矩母函数 $M(t)$ 的 n 阶导数为

$$M^n(t)=E[X^n e^{tX}] \qquad n\geqslant 1$$

由此得

$$M^n(0)=E[X^n] \qquad n\geqslant 1$$

下面对一些常见的分布函数求矩母函数.

例 6a（参数为 n 和 p 的二项分布） 设 X 是以 n 和 p 为参数的二项随机变量，则

$$M(t)=E[e^{tX}]=\sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k}$$

$$=\sum_{k=0}^{n}\binom{n}{k}(pe^t)^k(1-p)^{n-k}=(pe^t+1-p)^n$$

362 其中最后一个等式是根据二项式定理得来的. 对上式微分得

$$M'(t)=n(pe^t+1-p)^{n-1}pe^t$$

故

$$E[X]=M'(0)=np$$

这和例 1c 中得到的结果是一致的. 对上式求二阶微分得

$$M''(t)=n(n-1)(pe^t+1-p)^{n-2}(pe^t)^2+n(pe^t+1-p)^{n-1}pe^t$$

故

$$E[X^2]=M''(0)=n(n-1)p^2+np$$

X 的方差为

$$\mathrm{Var}(X)=E[X^2]-(E[X])^2=n(n-1)p^2+np-n^2p^2=np(1-p)$$

这也证实了例 3b 的结论. ■

例 6b（均值为 λ 的泊松分布） 如果 X 是以 λ 为参数的泊松随机变量，则

$$M(t)=E[e^{tX}]=\sum_{n=0}^{\infty}\frac{e^{tn}e^{-\lambda}\lambda^n}{n!}=e^{-\lambda}\sum_{n=0}^{\infty}\frac{(\lambda e^t)^n}{n!}=e^{-\lambda}e^{\lambda e^t}=\exp\{\lambda(e^t-1)\}$$

微分可得

$$M'(t)=\lambda e^t\exp\{\lambda(e^t-1)\}$$

$$M''(t)=(\lambda e^t)^2\exp\{\lambda(e^t-1)\}+\lambda e^t\exp\{\lambda(e^t-1)\}$$

因此

$$E[X]=M'(0)=\lambda$$

$$E[X^2]=M''(0)=\lambda^2+\lambda$$

$$\mathrm{Var}(X)=E[X^2]-(E[X])^2=\lambda$$

363 于是泊松分布的均值和方差均为 λ. ■

例 6c（参数为 λ 的指数分布） 假设 X 服从参数为 λ 的指数分布，则

$$M(t)=E[e^{tX}]=\int_0^{\infty} e^{tx}\lambda e^{-\lambda x}\,dx=\lambda\int_0^{\infty} e^{-(\lambda-t)x}\,dx=\frac{\lambda}{\lambda-t} \qquad (t<\lambda)$$

由上面的结果可知，对于指数分布，矩母函数 $M(t)$ 只在 $t<\lambda$ 时有定义. 对 $M(t)$ 微分可得

$$M'(t)=\frac{\lambda}{(\lambda-t)^2} \qquad M''(t)=\frac{2\lambda}{(\lambda-t)^3}$$

因此

$$E[X]=M'(0)=\frac{1}{\lambda} \qquad E[X^2]=M''(0)=\frac{2}{\lambda^2}$$

可知 X 的方差为

$$\mathrm{Var}(X)=E[X^2]-(E[X])^2=\frac{1}{\lambda^2}$$ ■

例 6d（正态分布） 首先计算单位正态分布 $N(0,1)$ 的矩母函数. 令 Z 为单位正态随机变量，我们有

$$\begin{aligned}M_Z(t) &= E[e^{tZ}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx} e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left\{-\frac{(x^2-2tx)}{2}\right\}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left\{-\frac{(x-t)^2}{2}+\frac{t^2}{2}\right\}dx \\ &= e^{t^2/2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(x-t)^2/2}dx = e^{t^2/2}\end{aligned}$$

364

因此单位正态随机变量 Z 的矩母函数 $M_Z(t)=e^{t^2/2}$. 为求得任一正态随机变量的矩母函数，我们回忆当 Z 为单位正态随机变量时，$X=\mu+\sigma Z$ 是以 μ 和 σ^2 为参数的正态随机变量（见 5.4 节）. 因此 X 的矩母函数可表示为

$$\begin{aligned}M_X(t) &= E[e^{tX}]=E[e^{t(\mu+\sigma Z)}]=E[e^{t\mu}e^{t\sigma Z}]=e^{t\mu}E[e^{t\sigma Z}] \\ &= e^{t\mu}M_Z(t\sigma)=e^{t\mu}e^{(t\sigma)^2/2}=\exp\left\{\frac{\sigma^2t^2}{2}+\mu t\right\}\end{aligned}$$

对上式微分可得

$$M'_X(t)=(\mu+t\sigma^2)\exp\left\{\frac{\sigma^2t^2}{2}+\mu t\right\}$$

$$M''_X(t)=(\mu+t\sigma^2)^2\exp\left\{\frac{\sigma^2t^2}{2}+\mu t\right\}+\sigma^2\exp\left\{\frac{\sigma^2t^2}{2}+\mu t\right\}$$

因此

$$E[X]=M'(0)=\mu$$
$$E[X^2]=M''(0)=\mu^2+\sigma^2$$

由此得

$$\mathrm{Var}(X)=E[X^2]-(E[X])^2=\sigma^2$$ ■

表 7-1 和表 7-2 给出了一些常见的离散型或连续型随机变量的矩母函数.

表 7-1 离散概率分布

	概率质量函数 $p(x)$	矩母函数 $M(t)$	均 值	方 差
参数为 n,p 的二项分布，其中 $0\leqslant p\leqslant 1$	$\binom{n}{x}p^x(1-p)^{n-x}$ $x=0,1,\cdots,n$	$(pe^t+1-p)^n$	np	$np(1-p)$
参数为 $\lambda>0$ 的泊松分布	$e^{-\lambda}\frac{\lambda^x}{x!}$ $x=0,1,2,\cdots$	$\exp\{\lambda(e^t-1)\}$	λ	λ
参数为 p 的几何分布，其中 $0\leqslant p\leqslant 1$	$p(1-p)^{x-1}$ $x=1,2,\cdots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
参数为 r,p 的负二项分布，其中 $0\leqslant p\leqslant 1$	$\binom{n-1}{r-1}p^r(1-p)^{n-r}$ $n=r,r+1,\cdots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

表 7-2 连续概率分布

	概率质量函数 $f(x)$	矩母函数 $M(t)$	均 值	方 差
(a,b) 上的一致分布	$f(x)=\begin{cases}\frac{1}{b-a} & a<x<b \\ 0 & 其他\end{cases}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
参数为 $\lambda>0$ 的指数分布	$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geqslant 0 \\ 0 & x<0\end{cases}$	$\frac{\lambda}{\lambda-t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
参数为 (s,λ) 的 Γ 分布，其中 $\lambda>0$	$f(x)=\begin{cases}\frac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)} & x\geqslant 0 \\ 0 & x<0\end{cases}$	$\left(\frac{\lambda}{\lambda-t}\right)^s$	$\frac{s}{\lambda}$	$\frac{s}{\lambda^2}$
参数为 (μ,σ^2) 的正态分布	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}\quad -\infty<x<\infty$	$\exp\left\{\mu t+\frac{\sigma^2 t^2}{2}\right\}$	μ	σ^2

矩母函数的一个重要性质是独立随机变量之和的矩母函数等于各个随机变量的矩母函数之积. 为证明这一事实，设随机变量 X 和 Y 相互独立，矩母函数分别为 $M_X(t)$ 和 $M_Y(t)$，则 $X+Y$ 的矩母函数 $M_{X+Y}(t)$ 为

$$M_{X+Y}(t)=E[e^{t(X+Y)}]=E[e^{tX}e^{tY}]=E[e^{tX}]E[e^{tY}]=M_X(t)M_Y(t)$$

其中倒数第二个等式是因为 X 和 Y 独立，用命题 3. 1 即得.

矩母函数的另一重要性质是矩母函数唯一决定分布函数. 也就是说，如果 $M_X(t)$ 在包含 $t=0$ 的某一区域内存在而且有限，则 X 的分布就被唯一确定. 例如，设 $M_X(t)=\left(\frac{1}{2}\right)^{10}(e^t+1)^{10}$，则由表 7-1 可知，$X$ 是以 10 和 $\frac{1}{2}$ 为参数的二项随机变量.

例 6c 设随机变量 X 的矩母函数为 $M(t)=e^{3(e^t-1)}$，求 $P\{X=0\}$.

解 由表 7-1 可知 $M(t)=e^{3(e^t-1)}$ 是均值为 3 的泊松随机变量的矩母函数，因此，根据矩母函数与分布函数之间的一一对应关系可知，X 一定是均值为 3 的泊松随机变量，于是 $P\{X=0\}=e^{-3}$ ■

例 6f（独立二项随机变量之和的分布） 如果 X 和 Y 分别服从参数为 (n,p) 和 (m,p) 的二项分布，且相互独立，求 $X+Y$ 的分布函数.

365～367

解 $X+Y$ 的矩母函数为

$$M_{X+Y}(t)=M_X(t)M_Y(t)=(pe^t+1-p)^n(pe^t+1-p)^m=(pe^t+1-p)^{m+n}$$

而 $(pe^t+1-p)^{m+n}$ 是参数为 $(m+n)$ 和 p 的二项分布的矩母函数，因此 $X+Y$ 服从参数为 $(m+n)$ 和 p 的二项分布. ■

例 6g（独立泊松随机变量之和的分布） 如果 X 和 Y 分别服从均值为 λ_1 和 λ_2 的泊松分布，且相互独立，求 $X+Y$ 的分布函数.

解

$$M_{X+Y}(t)=M_X(t)M_Y(t)=\exp\{\lambda_1(e^t-1)\}\exp\{\lambda_2(e^t-1)\}=\exp\{(\lambda_1+\lambda_2)(e^t-1)\}$$

因此 $X+Y$ 服从均值为 $\lambda_1+\lambda_2$ 的泊松分布，这也和第 6 章例 3d 的结论是一致的. ■

例 6h（独立正态随机变量之和的分布） 设 X 和 Y 分别服从参数为 (μ_1,σ_1^2) 和 (μ_2,σ_2^2) 的正态分布，且相互独立，则 $X+Y$ 服从参数为 $(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$ 的正态分布.

证明

$$M_{X+Y}(t)=M_X(t)M_Y(t)=\exp\left\{\frac{\sigma_1^2t^2}{2}+\mu_1t\right\}\exp\left\{\frac{\sigma_2^2t^2}{2}+\mu_2t\right\}$$
$$=\exp\left\{\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}+(\mu_1+\mu_2)t\right\}$$

这正是服从均值为$\mu_1+\mu_2$、方差为$\sigma_1^2+\sigma_2^2$的正态分布的矩母函数，由矩母函数与分布函数的唯一确定性可知结论成立. ■

例 6i 求自由度为n的卡方分布的矩母函数.

解 记$Z_1,\cdots,Z_n$为独立的标准正态随机变量，则卡方分布随机变量可表示为$Z_1^2+\cdots+Z_n^2$. 用$M(t)$表示其矩母函数，则$M(t)=(E[e^{tZ^2}])^n$，其中Z为标准正态随机变量. 现在，有

368

$$E[e^{tZ^2}]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{tx^2}e^{-x^2/2}dx$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x^2/2\sigma^2}dx \qquad 其中\ \sigma^2=(1-2t)^{-1}$$
$$=\sigma=(1-2t)^{-1/2}$$

倒数第二个等式用到了标准正态分布(均值为0,方差为σ^2)密度函数的积分为1. 因此，

$$M(t)=(1-2t)^{-n/2}$$ ■

例 6j(随机变量的随机数之和的矩母函数) 设$X_1,X_2,\cdots$为一个独立同分布的随机变量序列，N为只取非负整数值的随机变量，且与$X_i(i\geqslant1)$相互独立，求$Y=\sum_{i=1}^{N}X_i$的矩母函数.

在例4d中,当每个顾客消费的钱数与一天内进入商店的顾客数都是随机变量时，Y被解释为某天顾客在某商店消费的钱数。

解 为计算Y的矩母函数，先取N作为条件：

$$E\left[\exp\left\{t\sum_{i=1}^{N}X_i\right\}\middle|N=n\right]=E\left[\exp\left\{t\sum_{i=1}^{n}X_i\right\}\middle|N=n\right]=E\left[\exp\left\{t\sum_{i=1}^{n}X_i\right\}\right]=[M_X(t)]^n$$

这里

$$M_X(t)=E[e^{tX_i}]$$

因此

$$E[e^{tY}|N]=(M_X(t))^N$$

可得

$$M_Y(t)=E[(M_X(t))^N]$$

Y的各阶矩可以通过对矩母函数微分得到：

$$M'_Y(t)=E[N(M_X(t))^{N-1}M'_X(t)]$$

369

故

$$E[Y]=M'_Y(0)=E[N(M_X(0))^{N-1}M'_X(0)]=E[NE[X]]=E[N]E[X] \qquad (6\text{-}2)$$

这也验证了例4d的结论.(后面的式子用到了$M_X(0)=E[e^{0X}]=1$.)

同样，

$$M''_Y(t)=E[N(N-1)(M_X(t))^{N-2}(M'_X(t))^2+N(M_X(t))^{N-1}M''_X(t)]$$

故

$$E[Y^2]=M''_Y(0)=E[N(N-1)(E[X])^2+NE[X^2]]$$
$$=(E[X])^2(E[N^2]-E[N])+E[N]E[X^2]$$

$$=E[N](E[X^2]-(E[X])^2)+(E[X])^2E[N^2] \tag{6-3}$$
$$=E[N]\mathrm{Var}(X)+(E[X])^2E[N^2]$$

因此由式(6-2)和式(6-3)可知

$$\mathrm{Var}(Y)=E[N]\mathrm{Var}(X)+(E[X])^2(E[N^2]-(E[N])^2)$$
$$=E[N]\mathrm{Var}(X)+(E[X])^2\mathrm{Var}(N)$$

■

例 6k Y 服从(0,1)上的均匀分布，假设在给定 $Y=p$ 的条件下，随机变量 X 服从参数为 n 和 p 的二项分布. 在例 4j 中我们证明了 X 等可能地取 $0,1,\cdots,n$ 中的任何一个值，试用矩母函数的方法证明这一结论.

证明 为计算 X 的矩母函数，用 Y 的取值作为条件，由二项分布的矩母函数的公式可知

$$E[e^{tX}|Y=p]=(pe^t+1-p)^n$$

又由于 Y 服从(0,1)上的均匀分布，则有

$$E[e^{tX}]=\int_0^1(pe^t+1-p)^n\mathrm{d}p$$
$$=\frac{1}{e^t-1}\int_1^{e^t}y^n\mathrm{d}y \quad (\text{通过 } y=pe^t+1-p \text{ 变量替换})$$
$$=\frac{1}{n+1}\frac{e^{t(n+1)}-1}{e^t-1}=\frac{1}{n+1}(1+e^t+e^{2t}+\cdots+e^{nt})$$

370

由于上式为一个等可能地取 $0,1,\cdots,n$ 中的任何一个值的随机变量的矩母函数，所以由矩母函数与分布函数的唯一确定性可得结论成立. ■

联合矩母函数

我们也可以定义两个或更多个随机变量的联合矩母函数. 对任意 n 个随机变量 $X_1,X_2,\cdots,X_n$，它们的联合矩母函数 $M(t_1,\cdots,t_n)$ 可定义为：对任意实数 $t_1,\cdots,t_n$，

$$M(t_1,\cdots,t_n)=E[e^{t_1X_1+\cdots+t_nX_n}]$$

每一个随机变量 X_i 的矩母函数可以通过对 $M(t_1,\cdots,t_n)$ 的自变量部分令 t_i 以外的其他值取 0 而获得. 即

$$M_{X_i}(t)=E[e^{tX_i}]=M(0,\cdots,0,t,0,\cdots,0)$$

上式中 t 处于第 i 个位置.

同样可以证明 $M(t_1,\cdots,t_n)$ 可以唯一确定 $X_1,X_2,\cdots,X_n$ 的联合分布函数(该证明已超出本书内容). 这个结果可以用来证明 n 个随机变量 $X_1,\cdots,X_n$ 相互独立的充分必要条件是

$$M(t_1,\cdots,t_n)=M_{X_1}(t_1)\cdots M_{X_n}(t_n) \tag{6-4}$$

这是因为，如果 n 个随机变量是独立的，则

$$M(t_1,\cdots,t_n)=E[e^{(t_1X_1+\cdots+t_nX_n)}]=E[e^{t_1X_1}\cdots e^{t_nX_n}]$$
$$=E[e^{t_1X_1}]\cdots E[e^{t_nX_n}] \qquad \text{由独立性}$$
$$=M_{X_1}(t_1)\cdots M_{X_n}(t_n)$$

另一方面，如果式(6-4)成立，则联合矩母函数 $M(t_1,\cdots,t_n)$ 刚好等于 n 个独立随机变量的联合矩母函数，由联合矩母函数与联合分布函数的唯一确定性可知这 n 个随机变量是独立的.

例 6l 设 X 和 Y 为独立同分布的正态随机变量，均值为 μ、方差为 σ^2. 在第 6 章例 7a 中我们已经知道 $X+Y$ 和 $X-Y$ 也是独立的，下面我们通过联合矩母函数来证明这个结论.

证明

$$E[e^{t(X+Y)+s(X-Y)}]=E[e^{(t+s)X+(t-s)Y}]=E[e^{(t+s)X}]E[e^{(t-s)Y}]$$
$$=e^{\mu(t+s)+\sigma^2(t+s)^2/2}e^{\mu(t-s)+\sigma^2(t-s)^2/2}=e^{2\mu t+\sigma^2t^2}e^{\sigma^2s^2}$$

371

由上式我们知道这正是一个均值为 2μ、方差为 $2\sigma^2$ 的正态分布和一个与之独立的均值为 0、方差为 $2\sigma^2$ 的正态分布的联合矩母函数．由联合矩母函数与联合分布函数的唯一确定性可知，$X+Y$ 与 $X-Y$ 是相互独立的． ■

在下面的例子中我们将用联合矩母函数来证明第 6 章例 2b 中的结论．

例 6m 设事件发生的数目服从均值为 λ 的泊松分布，且每个事件被记录的概率为 p，则被记录的事件数与未被记录的事件数分别服从均值为 λp 和 $\lambda(1-p)$ 的泊松分布，且相互独立．

解 用 X 表示事件发生的总数目，X_c 表示被记录的事件数目，为计算被记录的事件数目 X_c 与未被记录的事件数目 $X-X_c$ 的联合矩母函数，以 X 的取值为条件得

$$E[e^{sX_c+t(X-X_c)}\mid X=n]=e^{tn}E[e^{(s-t)X_c}\mid X=n]=e^{tn}(pe^{s-t}+1-p)^n=(pe^s+(1-p)e^t)^n$$

上面的式子表明在 $X=n$ 的条件下，X_c 服从参数为 n 和 p 的二项分布．因此

$$E[e^{sX_c+t(X-X_c)}\mid X]=(pe^s+(1-p)e^t)^X$$

对上式两边取期望得

$$E[e^{sX_c+t(X-X_c)}]=E[(pe^s+(1-p)e^t)^X]$$

由于 X 服从均值为 λ 的泊松分布，$E[e^{tX}]=e^{\lambda(e^t-1)}$，故对任何正数 a（令 $a=e^t$）有，$E[a^X]=e^{\lambda(a-1)}$ 因此

$$E[e^{sX_c+t(X-X_c)}]=e^{\lambda(pe^s+(1-p)e^t-1)}=e^{\lambda p(e^s-1)}e^{\lambda(1-p)(e^t-1)}$$

由于上式是相互独立的分别服从均值为 λp 和 $\lambda(1-p)$ 的泊松分布的联合矩母函数，故原命题成立． ■

372

7.7 正态随机变量的其他性质

7.7.1 多元正态分布

令 $Z_1,\cdots,Z_n$ 为 n 个独立标准正态随机变量，如果对于正常数 $a_{ij}(1\leqslant i\leqslant m,1\leqslant j\leqslant n)$ 和 $\mu_i(1\leqslant i\leqslant m)$，

$$X_1=a_{11}Z_1+\cdots+a_{1n}Z_n+\mu_1$$
$$X_2=a_{21}Z_1+\cdots+a_{2n}Z_n+\mu_2$$
$$\vdots$$
$$X_i=a_{i1}Z_1+\cdots+a_{in}Z_n+\mu_i$$
$$\vdots$$
$$X_m=a_{m1}Z_1+\cdots+a_{mn}Z_n+\mu_m$$

则称随机变量族 $X_1,\cdots,X_m$ 服从多元正态分布．

由于独立正态随机变量之和仍服从正态分布，故 X_i 服从正态分布，其均值和方差为

$$E[X_i]=\mu_i$$

$$\mathrm{Var}(X_i)=\sum_{j=1}^{n}a_{ij}^2$$

我们考察 $X_1,\cdots,X_m$ 的联合矩母函数

$$M(t_1,\cdots,t_m)=E[\exp\{t_1X_1+\cdots+t_mX_m\}]$$

首先注意 $\sum_{i=1}^{m}t_iX_i$ 也可表示为独立正态分布 $Z_1,\cdots,Z_n$ 的线性组合，故也服从正态分布，其均

值和方差为

$$E\left[\sum_{i=1}^{m} t_i X_i\right] = \sum_{i=1}^{m} t_i \mu_i$$

$$\mathrm{Var}\left(\sum_{i=1}^{m} t_i X_i\right) = \mathrm{Cov}\left(\sum_{i=1}^{m} t_i X_i, \sum_{j=1}^{m} t_j X_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{m} t_i t_j \mathrm{Cov}(X_i, X_j)$$

如果 Y 为一均值为 μ、方差为 σ^2 的正态随机变量，则

373

$$E[\mathrm{e}^Y] = M_Y(t)\,|_{t=1} = \mathrm{e}^{\mu+\sigma^2/2}$$

因此可得

$$M(t_1, \cdots, t_m) = \exp\left\{\sum_{i=1}^{m} t_i \mu_i + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} t_i t_j \mathrm{Cov}(X_i, X_j)\right\}$$

由上可知 $X_1, \cdots, X_m$ 的联合分布可由各自的均值 $E[X_i]$ 和协方差 $\mathrm{Cov}(X_i, X_j)(i,j=1,\cdots,m)$ 完全确定.

7.7.2 样本均值和样本方差的联合分布

令 $X_1, \cdots, X_n$ 为独立同分布正态随机变量，均值为 μ，方差为 σ^2. 用 $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ 表示样本均值，由于独立正态随机变量之和仍服从正态分布，因此 $\overline{X}$ 服从均值为 μ、方差为 $\frac{\sigma^2}{n}$ 的正态分布(由例 2c 和例 3a)，

现在，回顾一下例 3f 的结论：

$$\mathrm{Cov}(\overline{X}, X_i - \overline{X}) = 0 \quad i = 1, \cdots, n \tag{7-1}$$

同样，$\overline{X}, X_1 - \overline{X}, \cdots, X_n - \overline{X}$ 均可表示为独立标准正态随机变量 $\frac{(X_i - \mu)}{\sigma}(i=1,\cdots,n)$ 的线性组合，即 $\overline{X}, X_i - \overline{X}(i=1,\cdots,n)$ 的联合分布为多元正态分布. 如果令 Y 为均值为 μ、方差为 $\frac{\sigma^2}{n}$ 的正态随机变量，且与 $X_i(i=1,\cdots,n)$ 相互独立，则 $Y, X_i - \overline{X}(i=1,\cdots,n)$ 的联合分布也为多元正态分布，由式(7-1)，与 $\overline{X}, X_i - \overline{X}(i=1,\cdots,n)$ 具有相同的均值和协方差. 但由于多元正态分布完全由各随机变量的均值和协方差确定，故 $Y, X_i - \overline{X}(i=1,\cdots,n)$ 与 $\overline{X}, X_i - \overline{X}(i=1,\cdots,n)$ 具有相同的联合分布，由此可得 $\overline{X}$ 与 $X_i - \overline{X}(i=1,\cdots,n)$ 是相互独立的.

由于 $\overline{X}$ 与 $X_i - \overline{X}(i=1,\cdots,n)$ 是相互独立的，故也与样本方差 $S^2 \equiv \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{(n-1)}$ 是独立的.

由于已知 $\overline{X}$ 服从均值为 μ、方差为 $\frac{\sigma^2}{n}$ 的正态分布，因此只需确定 S^2 的分布. 为此，由例 3a 中的等式

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\overline{X} - \mu)^2$$

上式两边同除以 σ^2 可得

374

$$\frac{(n-1)S^2}{\sigma^2} + \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \tag{7-2}$$

由于

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$$

是 n 个独立标准正态随机变量的平方和，故服从自由度为 n 的卡方分布，因此由例 6i 得，其矩母函数为 $(1-2t)^{-n/2}$. 同样 $\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$ 也是标准正态随机变量的平方，故服从自由度为 1 的卡方分布，矩母函数为 $(1-2t)^{-1/2}$. 前面我们已经知道式(7-2)的左边两项是独立的，因此由独立随机变量和的矩母函数等于各随机变量的矩母函数之积得到

$$E[e^{t(n-1)S^2/\sigma^2}](1-2t)^{-1/2}=(1-2t)^{-n/2}$$

或

$$E[e^{t(n-1)S^2/\sigma^2}]=(1-2t)^{-(n-1)/2}$$

而 $(1-2t)^{-(n-1)/2}$ 是自由度为 $n-1$ 的卡方分布的矩母函数，由于矩母函数唯一确定分布函数，故我们可得出结论，即 $\frac{(n-1)S^2}{\sigma^2}$ 服从自由度为 $n-1$ 的卡方分布.

综合以上我们得出下面的命题.

命题 7.1 *如果 $X_1,\cdots,X_n$ 为独立同分布的随机变量，且服从均值为 μ、方差为 σ^2 的正态分布，则样本均值 $\overline{X}$ 和样本方差 S^2 也是独立的，且 $\overline{X}$ 服从均值为 μ、方差为 σ^2/n 的正态分布，$\frac{(n-1)S^2}{\sigma^2}$ 服从自由度为 $n-1$ 的卡方分布.*

7.8 数学期望的一般定义

到目前为止，我们只对离散型和连续型随机变量定义了数学期望. 但除上述两种随机变量外，还存在既非连续型也非离散型的随机变量，这种随机变量也可以有数学期望. 下面举一个例子. 设 X 是以 $p=\frac{1}{2}$ 为参数的伯努利随机变量，Y 是在[0,1]上均匀分布的随机变量，并设 X 与 Y 独立，定义新的随机变量 W 为

$$W=\begin{cases}X & \text{如果 } X=1\\ Y & \text{如果 } X\neq 1\end{cases}$$

显然，W 既不是离散型（由于它的取值区间[0，1]是不可数的）也不是连续型 $\left(\text{由于 } P\{W=1\}=\frac{1}{2}\right)$ 的随机变量.

为了定义任意类型随机变量的数学期望，我们需要引入斯蒂尔切斯积分，在定义斯蒂尔切斯积分之前，回忆一下对函数 g，积分 $\int_a^b g(x)\mathrm{d}x$ 的定义为

$$\int_a^b g(x)\mathrm{d}x = \lim\sum_{i=1}^n g(x_i)(x_i-x_{i-1})$$

其中，极限取遍一切使得当 $n\to\infty$ 和 $\max\limits_{i=1,\cdots,n}(x_i-x_{i-1})\to 0$ 时的 $a=x_0<x_1<x_2<\cdots<x_n=b$.

对任意分布函数 F，我们定义非负函数 g 在区间 $[a,b]$ 上的斯蒂尔切斯积分为

$$\int_a^b g(x)\mathrm{d}F(x) = \lim\sum_{i=1}^n g(x_i)[F(x_i)-F(x_{i-1})]$$

如前面一样，其中极限取遍一切使得当 $n\to\infty$ 和 $\max\limits_{i=1,\cdots,n}(x_i-x_{i-1})\to 0$ 时的 $a=x_0<x_1<\cdots<x_n=b$. 进一步，我们在整个实数轴上定义斯蒂尔切斯积分为

$$\int_{-\infty}^{\infty} g(x)\mathrm{d}F(x) = \lim_{\substack{a\to-\infty\\ b\to+\infty}}\int_a^b g(x)\mathrm{d}F(x)$$

375

最后，如果 $g(x)$不是非负函数，我们定义 g^-，g^- 为

$$g^-(x)=\begin{cases} g(x) & \text{如果 } g(x)\geqslant 0 \\ 0 & \text{如果 } g(x)<0 \end{cases}$$

$$g^-(x)=\begin{cases} 0 & \text{如果 } g(x)\geqslant 0 \\ -g(x) & \text{如果 } g(x)<0 \end{cases}$$

由于 $g(x)=g^+(x)-g^-(x)$，又 g^-，g^- 均为非负函数，于是定义

$$\int_{-\infty}^{+\infty} g(x)\mathrm{d}F(x) = \int_{-\infty}^{+\infty} g^-(x)\mathrm{d}F(x) - \int_{-\infty}^{+\infty} g^-(x)\mathrm{d}F(x)$$

并且如果 $\int_{-\infty}^{+\infty} g^+(x)\mathrm{d}F(x)$和 $\int_{-\infty}^{+\infty} g^-(x)\mathrm{d}F(x)$不都等于$+\infty$，就称 $\int_{-\infty}^{+\infty} g(x)\mathrm{d}F(x)$存在.

如果任意的随机变量 X 具有累积分布函数 F，我们定义其数学期望为

$$E[X]=\int_{-\infty}^{+\infty} x\mathrm{d}F(x) \tag{8-1}$$

376

可以证明，若 X 为离散型随机变量，且具有概率质量函数 $p(x)$，则

$$\int_{-\infty}^{+\infty} x\mathrm{d}F(x)=\sum_{x:p(x)>0} xp(x)$$

如果 X 为连续型随机变量，且具有密度函数 $f(x)$，则

$$\int_{-\infty}^{+\infty} x\mathrm{d}F(x)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$$

考虑到 $E[x]$有渐近和式

$$\sum_{i=1}^{n} x_i[F(x_i)-F(x_{i-1})]$$

读者会看到式(8-1)给出的 $E[x]$的直观意义. 由于 $F(x_i)-F(x_{i-1})$恰好等于 X 落入区间$(x_{i-1},x_i]$中的概率，所以这个渐近和式就是 X 在区间$(x_{i-1},x_i]$的近似值 x_i 乘上 X 落入这个区间的概率，再对所有的区间求和. 显然，当这些区间长度越来越小时，我们就得到 X 的"期望值".

斯蒂尔切斯积分的作用主要是在理论方面，因为它提供了一个定义和研究数学期望性质的简洁方法. 例如，在讨论内容相同的定理时，利用斯蒂尔切斯积分就可以避免分离散型和连续型两种情形去陈述和证明. 由于斯蒂尔切斯积分的性质和普通积分的性质有许多是相同的，所以本章中的所有证明都可以非常方便地用在一般场合.

小结

如果随机变量 X 与 Y 具有联合概率质量函数 $p(x,y)$，则

$$E[g(X,Y)]=\sum_y\sum_x g(x,y)p(x,y)$$

如果随机变量 X 与 Y 具有联合密度函数 $f(x,y)$，则

$$E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$$

由上述定义可得

$$E[X+Y]=E[X]+E[Y]$$

进一步推广得

$$E\left[\sum_{i=1}^{n} X_i\right]=\sum_{i=1}^{n} E[X_i]$$

随机变量 X 和 Y 的协方差为

$$\mathrm{Cov}(X,Y)=E[(X-E[X])(Y-E[Y])]=E[XY]-E[X]E[Y]$$

377

一个重要的恒等式为

$$\mathrm{Cov}\left(\sum_{i=1}^{n} X_i,\sum_{j=1}^{m} Y_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{Cov}(X_i,Y_j)$$

当 $n=m$ 时，令 $Y_i=X_i, i=1,\cdots,n$，可得

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n}\mathrm{Var}(X_i)+2\sum_{i<j}\sum\mathrm{Cov}(X_i,Y_j)$$

随机变量 X 和 Y 的相关系数 $\rho(X,Y)$ 定义为

$$\rho(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

如果 X 和 Y 均为离散型随机变量，则在 $Y=y$ 条件下 X 的条件期望为

$$E[X|Y=y]=\sum_{x} xP\{X=x|Y=y\}$$

如果 X 和 Y 均为连续型随机变量，则

$$E[X|Y=y]=\int_{-\infty}^{+\infty} xf_{X|Y}(x|y)$$

其中 $f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$ 为在 $Y=y$ 条件下 X 的条件概率密度函数．条件期望和普通的数学期望具有类似的性质，不同之处在于所有的概率建立在 $Y=y$ 的条件之上．

令 $E[X|Y]$ 为 Y 的函数，在 $Y=y$ 时取值为 $E[X|Y=y]$．一个非常有用的恒等式为

$$E[X]=E[E[X|Y]]$$

在离散型随机变量情况下，上面的等式表示为

$$E[X]=\sum_{y} E[X|Y=y]P\{Y=y\}$$

在连续型随机变量情况下，上面的等式表示为

$$E[X]=\int_{-\infty}^{+\infty} E[X|Y=y]f_Y(y)$$

上面的等式常应用于以某个随机变量 Y 的取值为条件来求 $E[X]$．另外，对任何事件 A，有 $P(A)=E[I_A]$，这里如果事件 A 发生则 $I_A=1$，否则 $I_A=0$，我们可以用这种方法来计算概率问题．

378

在 $Y=y$ 条件下 X 的条件方差定义为

$$\mathrm{Var}(X|Y=y)=E[(X-E[X|Y=y])^2|Y=y]$$

令 $\mathrm{Var}(X|Y)$ 是 Y 的函数，其在 $Y=y$ 时的取值为 $\mathrm{Var}(X|Y=y)$．下面就是有名的条件方差公式：

$$\mathrm{Var}(X)=E[\mathrm{Var}(X|Y)]+\mathrm{Var}(E[X|Y])$$

假设随机变量 X 的取值可以观察到，由此需要预测随机变量 Y 的取值．在这种情况下，可以证明在所有的预测中，$E[Y|X]$ 是与 Y 的均方误差最小的预测．

随机变量 X 的矩母函数定义为 $M(t)=E[e^{tX}]$，而 X 的各阶矩可以通过对矩母函数微分并在 $t=0$ 处取值而得到，即

$$E[X^n]=\frac{\mathrm{d}^n}{\mathrm{d}t^n}M(t)\Big|_{t=0}\qquad n=1,2,\cdots$$

关于矩母函数还有两个重要的性质，其一，随机变量的矩母函数唯一地确定其分布函数；其二，独立随机变量和的矩母函数等于各个随机变量矩母函数的积. 利用这些结论可以很容易得到独立的正态(泊松或Γ)分布随机变量之和仍服从正态(泊松或Γ)分布的简单证明.

如果$X_1, X_2, \cdots, X_m$均为有限个独立标准正态分布的线性组合，则称$X_1, X_2, \cdots, X_m$具有多元正态分布. 其联合分布由$E[X_i]$与$\text{Cov}(X_i, X_j)$来确定，其中$i, j=1,2,\cdots,m$.

如果随机变量$X_1, X_2, \cdots, X_n$为独立同分布的正态随机变量，则其样本均值$\overline{X}=\sum_{i=1}^{n}\frac{X_i}{n}$和样本方差$S^2=\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{n-1}$相互独立. 样本均值$\overline{X}$服从均值为$\mu$、方差为$\sigma^2/n$的正态分布；$(n-1)S^2/\sigma^2$服从自由度为$n-1$的卡方分布.

习题

1. 一名游戏者向空中掷一个骰子并同时抛一枚均匀硬币. 如果硬币正面朝上，则他将赢得所掷骰子点数两倍的钱数，如果硬币反面朝上，则他将赢得所掷骰子点数一半的钱数，试确定所得钱数的期望值.

379

2. "线索"游戏包含6名嫌疑犯、6件武器和9个房间，随机地从中选择每一项中的一项，游戏的目标是猜测被选中的三项.
   (a) 有多少种可能的解答方案？
   这种游戏的一种说法是：当每一个游戏者做出选择后，随机地从剩余的卡片中得到3张. 令S,W和R分别表示3张卡片为元素这样的集合中嫌疑犯、武器和房间的数目，且令X表示游戏者在观看自己的3张卡片后所有可能的解答方案数.
   (b) 用S,W和R来表示X.
   (c) 求$E[X]$.
3. 如果X和Y为服从$(0,1)$上均匀分布的独立随机变量，证明
$$E[|X-Y|^{\alpha}]=\frac{2}{(\alpha+1)(\alpha+2)} \quad \text{对于 } \alpha>0$$
4. 如果X和Y为独立随机变量，均等可能地取$1,2,\cdots,m$中的任何值，证明
$$E[|X-Y|]=\frac{(m-1)(m+1)}{3m}$$
5. 县医院位于一个3英里范围内的方形区域的中央，如果有事故发生在这个区域的内部，医院将派救护车去救援，行驶路线均为矩形的，记医院所在位置的坐标为$(0,0)$，则到事故发生地(x,y)的行驶路程为$|x|+|y|$. 如果事故发生的位置服从这个方形区域上的均匀分布，求救护车到达事故点所需行驶路程的期望值.
6. 一个骰子随机掷10次，求得到的总点数的期望值.
7. 假设A和B随机独立地从10件不同的物品中选出3件，求下面的物品数目的期望值.
   (a) 被A和B均选中. (b) 未被A或B选中. (c) 被A或B中的任何一人选中.
8. N个人各自独立地进入一个餐厅，当到达餐厅时，每人都会观察一下是否有他的朋友在用餐. 如果有他的朋友在用餐，则他会坐在朋友的桌上；如果没有朋友在用餐，他会坐在一个空桌上. 假设所有$\binom{N}{2}$对人成为朋友的概率均为p，且相互独立，求被占用的餐桌数的期望值.
   提示：令X_i等于1还是0依赖于第i个人到来后是否坐在一个空桌上.
9. 有n个球，分别标号为从1到n，投向n个不同的坛子中，球$i(1\leqslant i\leqslant n)$等可能地被投入$n$个不同坛子中的任何一个.

(a) 求没有一个球投人的空坛子数目的期望值.

(b) 求所有坛子中都有球的概率.

10. 对于 3 次独立试验，每次成功的概率相同，用 Y 表示这些试验中成功的次数，如果 $E[X]=1.8$，求

(a) $P\{X=3\}$的最大可能取值.

(b) $P\{X=3\}$的最小可能取值.

在以上情况中，构造一种概率场景，使得 $P[X=3]$有规定的值.

提示：对于(b)，可以令 U 表示服从(0,1)上均匀分布的随机变量，然后根据 U 的值定义试验.

380

11. 假设独立地抛硬币 n 次，每次正面朝上的概率为 p. 当前后两次出现的正反面不同时，我们称为一次转换. 例如，当 $n=5$ 时，出现结果 $HHTHT$，则我们称有 3 次转换，求转换次数的期望值.

提示：将转换次数表示为 $n-1$ 个伯努利随机变量之和.

12. 有 n 名男士和 n 名女士随机地排为一队.

(a) 求有一名女士相邻的男士数的期望值.

(b) 当为随机环形排列时，求女士相邻的男士数的期望值.

13. 1 000 张分别标号为 1～1 000 的卡片，分发给 1 000 个人，求那些卡片标号和被分发人的年龄刚好相等的数目的期望值.

14. 坛子中有 m 个黑球，每次调换时将一个黑球取出，并放入一个新球，新球为黑球的概率为 p，为白球的概率为 $1-p$. 当坛子中不再有黑球时，求需要的调换次数的期望值.

注：上面的模型在理解艾滋病(AIDS)中有重要作用，人体内的免疫系统由一系列特定的细胞组成，例如 T 细胞. 这里有两种不同类型的 T 细胞，称为 CD4 和 CD8. 虽然 AIDS 患者的 T 细胞总数(至少在患病初期)与健康人相同，但最近已发现 CD4 型 T 细胞和 CD8 型 T 细胞的混合比例不同. 对于健康人来说，T 细胞中的 60%是 CD4 型，而对于艾滋病患者来说 CD4 型 T 细胞的比例将会持续下降. 一个新的模型提出 HIV 病毒(引发艾滋病的病毒)专门吞噬 CD4 型 T 细胞，而人体自身生成的新 T 细胞的各类数目不因被吞噬掉的是 CD4 型还是 CD8 型而不同，生成的新 T 细胞中以 0.6 的概率是 CD4 型，以 0.4 的概率是 CD8 型. 这样当人体内的 T 细胞等可能地被吞噬(以 0.6 的概率是 CD4 型)后，将逐步被新生成的 T 细胞替代，而如果病毒只吞噬 CD4 型 T 细胞，那将是一件很严重的事情.

4915. 在例 2h 中，对于 i 和 $j(i\neq j)$，如果第 i 个人取到的是第 j 个人的帽子，而第 j 个人取到的是第 i 个人的帽子，则称为一个配对，求配对数的期望值.

16. 令 Z 表示标准正态随机变量，对于固定的实数 x，定义

$$X=\begin{cases} Z & \text{如果 } Z>x \\ 0 & \text{否则} \end{cases}$$

试证明 $E[X]=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

17. 一套卡片共有 n 张，分别标号为 1 到 n，充分混合后排在一起，即 $n!$ 种可能的组合是等可能的. 假设我们依次进行 n 次猜测，第 i 次猜测是对第 i 个位置上卡片的猜测. 用 N 表示猜测正确的卡片数目.

(a) 如果在每次猜测时对以前的猜测正确与否全然不知，则在任何一种策略下，$E[N]=1$ 恒成立.

381

(b) 如果在每次猜测后均告知该位置的卡片是哪一张，则此时的最优策略是什么？在这种策略下证明

$$E[N]=\frac{1}{n}+\frac{1}{n-1}+\cdots+1\approx\int_1^n\frac{1}{x}dx=\log n$$

(c) 如果在每次猜测后均告知此次猜测正确与否，则证明此时的最优策略(使 $E[N]$最大)为：一直猜测同一张卡片直到正确为止，然后继续一直猜测另一张卡片. 在这种策略下，证明

$$E[N]=1+\frac{1}{2!}+\frac{1}{3!}+\cdots+\frac{1}{n!}\approx e-1$$

提示：在每种情况下，均可将 N 表示为示性(即伯努利型)随机变量和的形式.

18. 一副纸牌共 52 张，随机叠放在一起，每次翻开其中的一张. 如果第一张翻开的为 A，或者第二张为 2 点，或者第三张为 3 点，……，或者第十三张为 13 点，或者第十四张为 A，等等，我们就称出现一次匹配. 对于一次匹配而言，并不要求第$(13n+1)$张纸牌是特定花色的 A，而只要求是 A 即可，求翻遍一副纸牌所出现的匹配数目的期望值.

19. 某区域有 r 种不同类型的昆虫，且每种类型的昆虫被扑杀的概率为，$P_i(i=1,2,\cdots,r)$，$\sum_{i=1}^{r} P_i=1$，且对每种类型昆虫的扑杀是独立的.
 (a) 求在第 1 种类型昆虫被扑杀之前被扑杀的昆虫数目的期望数.
 (b) 求在第 1 种类型昆虫被扑杀之前被扑杀的昆虫类型的期望数.

20. 一个坛子中有 n 个球，第 i 个球的权重为 $W(i)$，$i=1,2,\cdots,n$. 每次从中按如下规则无放回地取球：每次取球时，任何一个球被选中的概率等于其权重与坛子中剩余球的总权重的比值. 例如，在某一时刻坛子中剩余的球为 $i_1,i_2,\cdots,i_r$，则下一次选中球 i_j 的概率为 $W(i_j)/\sum_{k=1}^{r} W(i_k)$，$j=1,2,\cdots,r$. 试计算在球 1 被取走之前已被取走的球数的期望值.

21. 对某 100 人而言，求(a)一年中恰有 3 个人在同一天过生日的天数的期望值；(b) 生日不同的人数的期望值.

22. 在一个均匀骰子的所有 6 面都至少出现一次之前，你平均要掷多少次?

23. 箱 I 中装有 5 个白球和 6 个黑球，箱 II 中装有 8 个白球和 10 个黑球. 现从箱 I 中随机取出两个球放入箱 II，然后从箱 II 中随机取出 3 个球，试求 3 个球中白球数的期望值.

382

提示：如果最初在箱 I 中的第 i 个白球是三个被取走的球之一，令 $X_i=1$；否则，令 $X_i=0$. 类似地，如果箱 II 中的第 i 个白球是三个被取走的球之一，令 $Y_i=1$；否则，令 $Y_i=0$. 于是，在三个被取走的球中，白球的个数等于 $\sum_{i=1}^{5} X_i+\sum_{i=1}^{8} Y_i$.

24. 一个药瓶中装有 m 粒大药片和 n 粒小药片，一个病人每天随机地从中取出一粒服用，如果取出的是小药片，则完全服用；如果取出的是大药片，则服用药片的一半，另一半放回药瓶中并视为一粒小药片.
 (a) 用 X 表示在大药片全部用完且将所有的另一半放回后药瓶中剩余的小药片数目，求 $E[X]$.
 提示：可定义 $n+m$ 个示性随机变量，分别表示原来的小药片与由大药片所产生的小药片，然后利用例 2 m 中的方法.
 (b) 用 Y 表示当最后一粒大药片用完时的天数，求 $E[Y]$.
 提示：利用 X 和 Y 的关系.

25. $X_1,X_2,\cdots$为独立同分布的连续型随机变量序列，令 $N\geqslant 2$ 且满足

$$X_1\geqslant X_2\geqslant\cdots\geqslant X_{N-1}<X_N$$

即 N 表示序列值不再增大的那一个变量序号，证明 $E[N]=\mathrm{e}$.
提示：首先计算 $P\{N\geqslant n\}$.

26. 设 $X_1,X_2,\cdots,X_n$ 为服从$(0,1)$上均匀分布的独立同分布随机变量.
 (a) 求 $E[\max(X_1,\cdots,X_n)]$. (b) 求 $E[\min(X_1,\cdots,X_n)]$.

*27. 如果有 101 个产品随机分装在 10 个箱子中，则至少有一个箱子中装的产品不少于 10 个，试用概率方法证明之.

*28. k-r-$n(k\leqslant r\leqslant n)$环形可靠性系统包括 n 个组件，且这 n 个组件组成环形. 每一个组见要么工作要么失效，并且如果每 r 个连续的组件中少于k 个失效，那么系统正常工作. 试证明没有一种方法安排 47 个组件，其中 8 个失效而使得 3-12-47 环形可靠性系统正常工作.

*29. 有四种不同类型的票券，前两种组成一组，后两种组成另一组. 获得第 i 种票券的概率为 p_i，其中 $p_1=p_2=1/8$，$p_3=p_4=3/8$. 试求至少得到下列之一所需要的票券数的期望值.

(a) 所有 4 种票券；(b) 第一组的两种票券；(c) 第二组的两种票券；(d) 任何一组的两种票券.

30. 如果 X 和 Y 为独立同分布的随机变量，且均值为 μ、方差为 σ^2，求 $E[(X-Y)^2]$.

31. 在习题 6 中，计算所掷的点数和的方差.

32. 在习题 9 中，计算空坛子数目的方差.

383

33. 如果 $E[X]=1$ 且 $\mathrm{Var}(X)=5$，求(a)$E[(2+X)^2]$；(b) $\mathrm{Var}(4+3X)$.

34. 设 10 对夫妇随机地围着圆桌坐下，求(a) 坐在自己丈夫身边的妻子人数的数学期望；(b) 坐在自己丈夫身边的妻子人数的方差.

35. 一副纸牌，每次从中翻开一张，为了得到下列情况下的纸牌，试求需翻开纸牌张数的期望值.
(a) 2 张 A；(b) 5 张黑桃；(c) 全部 13 张红心.

36. 掷一个均匀骰子 n 次，设 X 表示 1 出现的次数，Y 表示 2 出现的次数，求 $\mathrm{Cov}(X,Y)$.

37. 掷一个骰子两次，设 X 表示出现的点数之和，Y 表示第一次出现的点数减去第二次出现的点数之差，求 $\mathrm{Cov}(X,Y)$.

38. 设随机变量 X 和 Y 具有以下联合密度函数：

$$f(x,y)=\begin{cases}2\mathrm{e}^{-2x}/x & 0\leqslant x<\infty,\ 0\leqslant y\leqslant x\\ 0 & \text{否则}\end{cases}$$

试求 $\mathrm{Cov}(X,Y)$.

39. 设 $X_1,X_2,\cdots$ 为独立随机变量，具有相同的均值 μ 和方差 σ^2，令 $Y_n=X_n+X_{n+1}+X_{n+2}$，对于任意 $j\geqslant 0$，求 $\mathrm{Cov}(Y_n,Y_{n-j})$.

40. 如果随机变量 X 和 Y 具有以下联合密度函数：

$$f(x,y)=\frac{1}{y}\mathrm{e}^{-(y-x/y)},x>0,y>0$$

求 $E[X]$，$E[Y]$，并证明 $\mathrm{Cov}(X,Y)=1$.

41. 一个池塘中有 100 条鱼，其中有 30 条是鲤鱼. 如果从池塘中捕捞 20 条鱼，则捕上来的 20 条鱼中鲤鱼数的数学期望和方差是多少？你采用了什么样的假设？

42. 将 20 个人(其中有 10 名男士和 10 名女士)随机地按两个一对分成 10 对，试计算刚好是一男一女组合的对数的期望值和方差. 又假设 20 个人为 10 对夫妇，求刚好夫妇二人分为一组的组数的均值和方差.

43. 令 $X_1,X_2,\cdots,X_n$ 为独立随机变量且具有未知的连续型分布函数 F，令 $Y_1,Y_2,\cdots,Y_n$ 为独立随机变量且具有未知的连续型分布函数 G. 现对这 $n+m$ 个随机变量的取值进行排序，并令

$$I_i=\begin{cases}1 & \text{如果 } n+m \text{ 个变量中第 } i \text{ 个最小的变量是来自 } X \text{ 的样本}\\ 0 & \text{其他}\end{cases}$$

随机变量 $R=\sum_{i=1}^{n+m} iI_i$ 表示 X 样本的秩和，是检验分布函数 F 和 G 是否同分布的一种标准统计方法(称为威尔科克森秩和检验)的基础，这个检验在当 R 既不很大也不很小时接受假设：$F=G$. 假定相等性的假设成立，试计算 R 的均值和方差.
提示：利用例 3c 的结果.

44. 有两种不同的工艺来加工某种产品，由第 i 种工艺加工的产品质量指标服从连续型分布函数 F_i，$i=1,2$. 假设有 n 件产品由第一种工艺加工，m 件产品由第二种工艺加工，将 $n+m$ 件产品按质量指标大小排序，令

$$X_j=\begin{cases}1 & \text{若第 } j \text{ 个最好的产品由第一种工艺加工}\\ 2 & \text{其他}\end{cases}$$

384

对于包括 n 个 1 和 m 个 2 的向量 $X_1,X_2,\cdots,X_{n+m}$，用 R 表示 1 的游程数. 例如，若 $n=5$，$m=2$，且 $X=1,2,1,1,1,1,2$，则 $R=2$. 如果 $F_1=F_2$，即两种方法加工的产品的质量指标具有相同的分布，求 R

的均值和方差.

45. 设 X_1, X_2, X_3, X_4 为不相关的随机变量，分别具有相同的均值 0 和方差 1，计算下列相关系数.
(a) X_1+X_2 和 X_2+X_3；(b) X_1+X_2 和 X_3+X_4

46. 在下面的掷骰子赌博游戏中，玩家 1 和玩家 2 轮流掷骰子，然后由庄家掷骰子以确定输赢. 对于玩家 i，$i=1,2$，如果他的点数严格大于庄家的点数则为赢. 对于 $i=1,2$，令

$$I_i=\begin{cases}1 & \text{如果玩家 } i \text{ 赢}\\ 0 & \text{否则}\end{cases}$$

试证明 I_1 和 I_2 是正相关的，并解释其中的原因.

47. 考虑一个有 n 个顶点的图，这些顶点分别记为 $1,2,\cdots,n$，假设 $\binom{n}{2}$ 对不同的顶点间的连线中的任一条都以概率 p 出现，且相互独立. 顶点 i 的次数是顶点 i 作为顶点之一的边的个数，记为 D_i.
(a) D_i 的分布是什么？
(b) 求 D_i 和 D_j 的相关系数 $\rho(D_i,D_j)$.

48. 连续掷一个均匀骰子，用 X 和 Y 分别表示为获得 6 点和 5 点所需要的掷骰子的次数，求(a) $E[X]$；(b) $E[X|Y=1]$；(c) $E[X|Y=5]$.

49. 坛子中装有 4 个白球和 6 个黑球，不放回地接连从坛子中取出容量分别为 3 和 5 的两个随机样本. 设 X 和 Y 分别表示这两个样本中的白球数，试对 $i=1,2,3,4$ 求 $E[X|Y=i]$.

50. 设随机变量 X 和 Y 的联合密度函数为

$$f(x,y)=\frac{e^{-x/y}e^{-y}}{y}\quad 0<x<\infty,0<y<\infty$$

求 $E[X^2|Y=y]$.

51. 设随机变量 X 和 Y 的联合密度函数为

$$f(x,y)=\frac{e^{-y}}{y}\quad 0<x<y,0<y<\infty$$

求 $E[X^3|Y=y]$.

52. 一个总集合由 r 个不同的子集构成，用 p_i 表示总集合中的元素归属第 i 个子集的概率，其中 $i=1,2,\cdots,r$. 如果第 i 个子集中元素的平均权重为 $w_i,i=1,2,\cdots,r$，则总集合中元素的平均权重为多少？

53. 一囚犯困在有 3 个门的密室中，第一个门通向一地道，沿此地道行走 2 天后，结果又转回原地；第二个门通向行走 4 天后也回到原地的地道；第三个门通向行走 1 天就能得到自由的地道. 设该囚犯始终以概率 0.5,0.3 和 0.2 分别选择第一个门、第二个门和第三个门，试问该囚犯走出地道获得自由平均需要多少天？

54. 考虑下面的掷骰子游戏，掷一对骰子，如果所得点数之和为 7，则游戏结束，且游戏者积分为 0. 如果所得点数之和不是 7，则此时游戏者可做如下的选择：或者停止游戏得到和本次所得点数之和相同的积分值，或者继续进行游戏. 对每一个取值 i，$i=1,2,\cdots,12$，如果你采用下面的策略：当所得点数之和大于或等于 i 时，就停止游戏，则所得积分的期望值是多少？i 取值为多少时能使积分期望值最大？
提示：用 X_i 表示所采用策略中设定的界值为 i 时得到的积分值，为计算 $E[X_i]$，可以对首次所得点数值取条件概率.

385

55. 有 10 名猎人正等着野鸭飞来，当一群野鸭飞过头顶时，10 人同时开枪射击，但他们是彼此独立地、随机地选择自己的目标. 假定每一名猎人独立地击中目标的概率为 0.6，试求被他们击中的野鸭数的期望值，其中假设这群野鸭数是均值为 6 的泊松随机变量.

56. 在底层乘电梯的人数是均值为 10 的泊松随机变量，设共有 N 层楼，又每一名乘客在哪一层下是等可能的，而且他到哪一层下与其余乘客在这一层下是彼此独立的，求在所有乘客都走出电梯之前，该电梯停的次数的期望值.

57. 设某工厂每周平均发生 5 次事故，且在每次事故中，受伤的工人人数是彼此独立的随机变量，具有相同的均值 2.5. 如果在每次事故中，受伤的工人人数与事故发生的次数是独立的，试求一周内受伤工人的平均人数.

58. 连续抛一枚正面出现的概率为 p 的硬币，直到正面和反面都出现过，求
(a) 需要抛的次数的期望值；(b) 最后一次抛出现正面的概率.

59. 一个人连续抛一枚硬币直到连续出现 3 次正面为止，假设每次抛是独立的且抛出正面的概率为 p. 试确定需要抛的次数的期望值.
提示：用 T 表示首次抛出现反面，并令全部抛出正面时 $T=0$，然后对 T 取条件概率.

60. 有 $n+1$ 个人参与某项游戏，每个人相互独立地成为赢家的概率为 p，所有赢家分享一单位的奖金（例如，有 4 个人成为赢家，则每人得到奖金的 $\frac{1}{4}$，而如果没有人成为赢家，则所有参与者不得任何奖金）. 令 A 表示某一参与者，X 表示 A 获得的奖金数目.
(a) 求参与者得到的奖金数目的期望值.
(b) 证明 $E[X]=\frac{1-(1-p)^{n+1}}{n+1}$.
(c) 以 A 是否是赢家为条件，计算 $E[X]$，并证明
$$E[(1+B)^{-1}]=\frac{1-(1-p)^{n+1}}{(n+1)p}$$
这里 B 是参数为 n 和 p 的二项随机变量.

61. 有 $m+2$ 个玩家进行抛均匀硬币的赌博游戏，每次下赌注 1 美元，每人连续抛 n 次，n 为奇数，并记录每次的结果. 在每个玩家抛硬币之前要对他将要得到的结果进行预测. 例如，当 $n=3$ 时，如果一人写下（H、H、T），即表示他预测第一次抛出正面，第二次抛出正面，第三次抛出反面. 抛完后将真实结果与预测结果进行比较，如果所抛结果全是正面，则其预测对了两个，所有的 $m+2$ 美元赌注将被那些预测正确数最多的玩家所均分.
由于每次抛硬币所得正反面是等概率的，前 m 个玩家将采用随机的预测办法，连续抛 n 次后使用结果作为预测. 而后两人将组成一个联合组并采用下面的策略：一个玩家将和其他 m 人一样随机地写下预测结果，而另一个玩家将会与其预测结果完全相反. 即如果一个人预测某一次结果为 H，则他将预测为 T. 例如，一个人写下预测结果为（H,H,T），则另一个人写下的预测结果将为（T、T、H）.
(a) 证明联合组的两个人中恰好有一人会有大于 $n/2$（n 为奇数）次的正确预测.
(b) 用 X 表示前 m 个非联合玩家中预测正确的次数大于 $n/2$ 的人数，X 的分布是什么？
(c) X 如(b)所定义，证明
$$E[\text{支付给联合组的钱数}]=(m+2)E\left[\frac{1}{X+1}\right]$$
(d) 利用习题 60 中(c)的结论证明
$$E[\text{支付给联合组的钱数}]=\frac{2(m+2)}{m+1}\left[1-\left(\frac{1}{2}\right)^{m+1}\right]$$
并写出当 $m=1,2$ 和 3 时的明确表达式. 由于可以证明
$$\frac{2(m+2)}{m+1}\left[1-\left(\frac{1}{2}\right)^{m+1}\right]>2$$
因此组成联合组的两个人的收益永远为正的.

62. 设 $U_1,U_2,\cdots$ 为服从 (0,1) 上均匀分布的独立随机变量，在例 4h 中我们已经得到对于 $0\leqslant x\leqslant 1$，有 $E[N(x)]=e^x$，这里 $N(x)=\min\left\{n:\sum_{i=1}^{n}U_i>x\right\}$，下面给出这一结论的另一种证明方法.
(a) 用归纳法证明，对于 $0\leqslant x\leqslant 1$ 和 $n\geqslant 0$，有

386

$$P\{N(x)\geqslant n+1\}=\frac{x^n}{n!}$$

提示：以 U_1 为条件并利用归纳假设.

(b) 利用(a)证明

$$E[N(x)]=e^x$$

63. 一个坛子中装有 30 个球，其中有 10 个红球、8 个蓝球，从这个坛子中随机取出 12 个球，用 X 表示被取出的红球个数，Y 表示被取出的蓝球个数，计算 $\text{Cov}(X,Y)$.

387

(a) 通过定义适当的示性随机变量(伯努利型) X_i,Y_j，使得

$$X=\sum_{i=1}^{10}X_i,Y=\sum_{j=1}^{8}Y_j$$

(b) 通过以 X 或 Y 为条件计算 $E[XY]$.

64. 第 i 类型灯泡的寿命是均值为 μ_i、标准差为 σ_i 的随机变量，其中 $i=1,2$. 现从一个大箱子中随机取出一只灯泡，它为第 1 类型的概率为 p，为第 2 类型的概率为 $1-p$，用 X 表示这只灯泡的寿命，求

(a) $E[X]$；　(b) $\text{Var}(X)$.

65. 在一个气候良好的年份里，冬季暴风雪天数为一均值为 3 的泊松随机变量，而在一个气候恶劣的年份里，冬季暴风雪天数为一均值为 5 的泊松随机变量. 如果下一年气候良好的概率为 0.4，气候恶劣的概率为 0.6，求下一年暴风雪天数的数学期望和方差.

66. 计算例 4c 中矿工安全脱险所需时间的方差.

67. 考虑一赌徒，在每一局赌博中，他赌赢或输掉赌注的概率分别为 p 和 $1-p$. 当 $p>\frac{1}{2}$时，一种常见的策略(称为 Kelley 策略)为：每次下的赌注等于你手中现有赌本的 $2p-1$ 倍，如果设开始时该赌徒的赌本为 x 美元，且他采用上述策略去赌博，求经过 n 局赌博后他的赌本的期望值.

68. 在给定的一年内，一人发生事故的次数是均值为 λ 的泊松随机变量. 设 λ 的值因人而异，λ 等于 2 的人占 60%，λ 等于 3 的人占 40%. 随机地挑选一人，问下列事件的概率是多少?

(a) 在一年内将发生 0 次事故；(b) 一年内恰好有 3 次事故发生.

如已知他在上一年没发生事故，则下一年内他将发生 3 次事故的条件概率是多少?

69. 在上题中如设均值 λ 小于 x 的人所占的比例为 $1-e^{-x}$，试重新计算上述概率.

70. 考虑一个装有大量硬币的箱子，并设抛其中一枚硬币时，出现正面的概率为 p. 但是，这个值随硬币而变. 假定箱子中的硬币具有这样的性质：若从中随机地取出一枚硬币，其 p 值可视为[0,1]上的均匀随机变量. 现设从箱子中随机取出一枚硬币并抛两次，求(a)第一次出现正面的概率；(b) 两次都出现正面的概率.

71. 在上题中，假设抛硬币 n 次，用 X 表示出现正面的次数，证明

$$P\{X=i\}=\frac{1}{n+1}\quad i=0,1,\cdots,n$$

提示：利用 $\int_0^1 x^{a-1}(1-x)^{b-1}dx=\frac{(a-1)!\ (b-1)!}{(a+b-1)!}$，其中 a,b 为正整数.

72. 假设在 70 题中我们连续抛硬币直到出现正面为止，用 N 表示需要抛的次数，求

(a) $P\{N\geqslant i\},i\geqslant 0$；(b) $P\{N=i\}$；(c) $E[N]$.

73. 在例 5b 中用 S 来表示发送的信号值，用 R 来表示接收到的信号值.

388

(a) 计算 $E[R]$. (b) 计算 $\text{Var}(R)$. (c) R 是否为正态分布? (d) 计算 $\text{Cov}(R,S)$.

74. 在例 5c 中，假设 X 服从(0,1)上的均匀分布，如果离散化后的值由 $a_0=0,a_1=\frac{1}{2},a_2=1$ 来确定，试确定最优的量化数值 Y 并计算 $E[(X-Y)^2]$.

75. 设随机变量 X 的矩母函数为 $M_X(t)=\exp(2e^t-2)$，而随机变量 Y 的矩母函数为 $M_Y(t)=\left(\frac{3}{4}e^t+\frac{1}{4}\right)^{10}$. 假如 X 与 Y 相互独立，求

(a) $P\{X+Y=2\}$；(b) $P\{XY=0\}$；(c) $E[XY]$.

76. 用 X 表示首次掷骰子出现的点数，Y 表示两次掷骰子的点数之和，计算 X 和 Y 的联合矩母函数.

77. 设随机变量 X 和 Y 的联合密度函数为

$$f(x,y)=\frac{1}{\sqrt{2\pi}}e^{-y}e^{-(x-y)^2/2}\quad 0<y<\infty,-\infty<x<\infty$$

(a) 计算 X 和 Y 的联合矩母函数.

(b) 计算 X 和 Y 各自的矩母函数.

78. 有两个信封，各装有一张支票，一个人将其中任何一个信封打开取出支票并看到支票上钱的数目. 现在他可以做两个选择，一是接受这个支票上的数目，二是和那个未打开信封里的支票相交换. 他将采用何种选择？是否用其他方案比直接接受更好一些？

不妨用 A 和 $B(A<B)$，来表示信封中的支票额(未知)，可知如果随机选择一个信封就接受其中的支票上的钱数的期望值为$\frac{A+B}{2}$. 考虑下面的策略：用 $F(\cdot)$表示一个连续递增的分布函数，任取一个信封打开，如果支票额为 x，则以概率 $F(x)$接受它，且以概率 $1-F(x)$拒绝它，然后交换另一个信封.

(a) 证明如果采用上面的策略，则得到的支票额的期望值会大于$\frac{A+B}{2}$.

提示：以第一次打开信封里的支票额为 A 或 B 为条件.

现在考虑下面这个策略：首先固定一个值 x，如果第一次打开信封得到支票额大于 x 就接受，否则就拒绝，然后选择另一个.

(b) 证明对于任何的 x 值，采用第二个策略所得到的支票额的期望值至少为$\frac{A+B}{2}$，并且如果 x 的值位于 A 和 B 之间，则所得到的支票额的期望值严格大于$\frac{A+B}{2}$.

(c) 令 X 为实数域上的连续型随机变量，考虑下面的策略：取 X 的一个值，如果这个值为 $X=x$，则采用上面的第二个策略并取固定值为 x，证明这种情况下所得支票额的期望值大于$\frac{A+B}{2}$.

理论练习

1. 证明 $E[(x-a)^2]$在 $a=E[X]$处取得最小值.

2. 假设 X 是密度函数为 f 的连续型随机变量. 证明 $E[|X-a|]$在 a 为 F 的中位数时取得最小值.

提示：写

$$E[|X-a|]=\int|x-a|f(x)\mathrm{d}x$$

将积分区域分为 $x>a$ 和 $x<a$ 两种情况考虑.

389

3. 在下列情况下证明命题 2.1：

(a) X 和 Y 具有联合质量函数.

(b) X 和 Y 具有联合密度函数 $g(x,y)$，且对所有 x,y 有 $g(x,y)\geqslant 0$.

4. 令 X 为一个具有有限均值 μ 和方差 σ^2 的随机变量，$g(\cdot)$为一个二阶可微函数，证明

$$E[g(X)]\approx g(\mu)+\frac{g''(\mu)}{2}\sigma^2$$

提示：将 $g(\cdot)$在 μ 处进行泰勒级数展开，然后取前三项而忽略其他项.

5. 令 $A_1, A_2, \cdots, A_n$ 表示一个事件序列，定义 $C_k=\{$至少有 k 个 A_i 事件发生$\}$，证明

$$\sum_{k=1}^{n} P(C_k)=\sum_{k=1}^{n} P(A_k)$$

提示：令 X 表示 A_i 的发生数，证明上式的两边均等于 $E[X]$.

6. 在本书中我们注意到

$$E\Big[\sum_{i=1}^{\infty} X_i\Big]=\sum_{i=1}^{\infty} E[X_i]$$

这里 X_i 全部为非负随机变量. 因为积分可以认为是和的极限，我们期望

$$E\Big[\int_0^{\infty} X(t)\mathrm{d}t\Big]=\int_0^{\infty} E[X(t)]\mathrm{d}t$$

这里 $X(t)(0\leqslant t<\infty)$，均为非负随机变量，这个结论确实成立. 利用这个结论证明，对于非负随机变量 X 有

$$E[X]=\int_0^{\infty} P\{X>t\}\mathrm{d}t$$

提示：对于每个非负值 t，定义随机变量 $X(t)$ 为

$$X(t)=\begin{cases}1 & \text{如果 } t<X\\ 0 & \text{如果 } t\geqslant X\end{cases}$$

并将 $\int_0^{\infty} X(t)\mathrm{d}t$ 和 X 联系起来.

7. 假设随机变量 X 和 Y 对所有实数 t，有 $P\{X>t\}\geqslant P\{Y>t\}$成立，则我们称 X 随机大于 Y，记作 $X\geqslant_{st}Y$. 对于下列情况，证明如果 $X\geqslant_{st}Y$，则 $E[X]\geqslant E[Y]$.

(a) X 和 Y 均为非负随机变量.

(b) X 和 Y 为任意的随机变量.

390

提示：将 X 表示为 $X=X^+-X^-$，这里，

$$X^+=\begin{cases}X & \text{如果 } X\geqslant 0\\ 0 & \text{如果 } X<0\end{cases},\quad X^-=\begin{cases}0 & \text{如果 } X\geqslant 0\\ -X & \text{如果 } X<0\end{cases}$$

同样，可将 Y 表示成 $Y=Y^+-Y^-$，然后利用(a)中的结论.

8. 证明随机变量 X 随机大于随机变量 Y 当且仅当对任意的递增函数 f 有

$$E[f(X)]\geqslant E[f(Y)]$$

提示：如果 $X\geqslant_{st}Y$，可以利用 $f(X)\geqslant_{st}f(Y)$ 和上题证明 $E[f(X)]\geqslant E[f(Y)]$. 另一方面，利用 $E[f(X)]\geqslant E[f(Y)]$等价于对所有的递增函数 f，有 $P\{X>t\}\geqslant P\{Y>t\}$，构造一个适当的递增函数 f.

9. 连续抛一枚硬币 n 次，正面朝上的概率为 p. 如果在两次反面之间有连续的正面出现 r 次，就称为一个长为 r 的游程，求长为 1, 2, $k(1\leqslant k\leqslant n)$的游程出现的次数的期望值.

10. 设 $X_1, X_2, \cdots, X_n$ 为独立同分布的正随机变量，对 $k\leqslant n$，求 $E\left[\dfrac{\sum_{i=1}^{k} X_i}{\sum_{i=1}^{n} X_i}\right]$.

11. 考虑 n 次独立试验，每次可能出现的 r 种结果的概率分别为 $P_1, P_2, \cdots, P_r$. 用 X 表示在 n 次试验中没有出现的结果的个数. 求 $E[X]$，并证明在所有的概率向量 $P_1, P_2, \cdots, P_r$ 中，$E[X]$在 $P_i=1/r(i=1, 2, \cdots, r)$时取得最小值.

12. 进行一系列独立试验，第 i 次试验成功的概率为 P_i，在 n 次试验中，计算(a) 成功次数的数学期望；(b)成功次数的方差. 题中假设的独立性对上述两个问题的求解有什么作用?

13. 设 $X_1, X_2, \cdots, X_n$ 为独立同分布的连续型随机变量，我们称在时刻 $j(j\leqslant n)$出现一次峰值，如果对一切 $1\leqslant i\leqslant j$ 有 $X_i\leqslant X_j$. 试证：

(a) E[峰值出现的次数]$=\sum_{j=1}^{n}1/j$；(b) Var(峰值出现的次数)$=\sum_{j=1}^{n}(j-1)/j^2$.

14. 证明在例 2j 中为获得每种类型至少一张的整套票券，需要收集的票券数的方差等于$\sum_{i=1}^{N-1}\frac{iN}{(N-i)^2}$，并说明当 N 很大时，该值接近于 $N^2\pi^2/6$.

15. 对于 n 次独立试验，第 i 次成功的概率为 P_i.

    (a) 计算在 n 次试验中成功次数的期望值 μ.

    (b) 对于确定的 μ 值，如何选择 $P_1,\cdots,P_n$ 的值使得成功次数的方差最大?

    (c) 对于确定的 μ 值，如何选择 $P_1,\cdots,P_n$ 的值使得成功次数的方差最小?

391

*16. 假设集合 $S=\{1,2,\cdots,n\}$ 中的元素均要被涂成红色或蓝色，$A_1,A_2,\cdots,A_r$ 为 S 的子集，证明有一种涂色办法可使至多 $\sum_{i=1}^{r}(1/2)^{|A_i|-1}$ 个子集中的元素具有相同的颜色(其中 $|A|$ 表示集合 A 中的元素数).

17. 假设随机变量 X_1 和 X_2 为独立随机变量且具有相同的均值 μ，而 $\mathrm{Var}(X_1)=\sigma_1^2$，$\mathrm{Var}(X_2)=\sigma_2^2$. 均值 μ 未知，用 X_1，X_2 的加权和来估计 μ，即常用 $\lambda X_1+(1-\lambda)X_2$ 来估计 μ，λ 如何取值才能使估计具有最小方差? 解释 λ 取值的合理性.

18. 在例 3h 中，我们证明了多项分布型随机变量 N_i 和 N_j 的协方差为 $-mP_iP_j$，是通过将 N_i 和 N_j 表示成示性变量的和而完成的. 这个结论也可以利用公式 $\mathrm{Var}(N_i+N_j)=\mathrm{Var}(N_i)+\mathrm{Var}(N_j)+2\mathrm{Cov}(N_i,N_j)$ 得到.

    (a) 求 N_i+N_j 的分布.

    (b) 利用上面的等式证明 $\mathrm{Cov}(N_i,N_j)=-mP_iP_j$.

19. 设 X 和 Y 为同分布的随机变量，但不一定独立，证明

$$\mathrm{Cov}(X+Y,X-Y)=0$$

20. *条件协方差公式*. 在随机变量 Z 取值已定的条件下，随机变量 X 和 Y 的条件协方差定义为

$$\mathrm{Cov}(X,Y|Z)\equiv E[(X-E[X|Z])(Y-E[Y|Z])|Z]$$

    (a) 证明

$$\mathrm{Cov}(X,Y|Z)=E[XY|Z]-E[X|Z]E[Y|Z]$$

    (b) 证明条件协方差公式

$$\mathrm{Cov}(X,Y)=E[\mathrm{Cov}(X,Y|Z)]+\mathrm{Cov}(E[X|Z],E[Y|Z])$$

    (c) 在(b)中，令 $X=Y$，给出条件方差公式.

21. 用 $X_{(i)}(i=1,\cdots,n)$ 表示一个服从(0,1)上均匀分布的随机变量集合的顺序统计量，且 $X_{(i)}$ 的密度函数为

$$f(x)=\frac{n!}{(i-1)!\,(n-i)!}x^{i-1}(1-x)^{n-i}\quad 0<x<1$$

    (a) 计算 $\mathrm{Var}(X_{(i)}),i=1,\cdots,n$.

    (b) i 为何值时能使 $\mathrm{Var}(X_{(i)})$ 取值最大和最小?

22. 如果 $Y=a+bX$，证明

$$\rho(X,Y)=\begin{cases}+1 & \text{如果 } b>0\\ -1 & \text{如果 } b<0\end{cases}$$

23. 设 Z 为标准正态随机变量，如果随机变量 Y 定义为 $Y=a+bZ+cZ^2$，证明

$$\rho(Y,Z)=\frac{b}{\sqrt{b^2+2c^2}}$$

392

24. 证明柯西一施瓦茨不等式

$$(E[XY])^2\leqslant E[X^2]E[Y^2]$$

提示：当且仅当 $Y=-tX$ 时，上式等号成立. 对任意 t，

$$0<E[(tX+Y)^2]=E[X^2]t^2+2E[XY]t+E[Y^2]$$

因此二次方程 $E[X^2]t^2+2E[XY]t+E[Y^2]=0$ 只有虚根，故可得其判别式必为负值.

25. 假设随机变量 X 与 Y 独立，证明在下列情况下，对所有的 y 有 $E[X|Y=y]=E[X]$ 成立.
 (a) X 和 Y 为离散型随机变量.
 (b) X 和 Y 为连续型随机变量.
26. 证明：
$$E[g(X)Y|X]=g(X)E[Y|X]$$
27. 证明：如果 $E[Y|X=x]=E[Y]$ 对所有 x 成立，则随机变量 X 和 Y 是不相关的，并举反例说明其逆命题不成立.
 提示：可利用 $E[XY]=E[XE[Y|X]]$.
28. 证明：$\mathrm{Cov}(X,E[Y|X])=\mathrm{Cov}(X,Y)$.
29. 设 $X_1,X_2,\cdots,X_n$ 为独立同分布的随机变量，求 $E[X_1|X_1+\cdots+X_n=x]$.
30. 考虑例 3h 中有关多项分布随机变量，利用条件期望计算 $E[N_iN_j]$，并证明例 3h 中给出的有关 $\mathrm{Cov}(N_i,N_j)$ 的公式.
31. 一个坛子中原有 b 个黑球和 w 个白球. 每次我们放入 r 个黑球，并从其中的 $b+w+r$ 个球中再随机地取出 r 个球，证明
$$E(\text{第 } t \text{ 次抽球后坛中的白球数})=\left(\frac{b+w}{b+w+r}\right)^t w$$
32. 假设在 $Y=y$ 条件下，随机变量 X_1 和 X_2 是均值为 y 的独立随机变量，证明
$$\mathrm{Cov}(X_1,X_2)=\mathrm{Var}(Y)$$
33. 连续抛一枚均匀硬币，正面朝上的概率为 p，求直到出现 r 次连续正面为止所需要抛硬币的期望数.
 提示：以首次出现反面的时刻为条件可得方程
$$E[X]=(1-p)\sum_{i=1}^{r}p^{i-1}(i+E[X])+(1-p)\sum_{i=r+1}^{\infty}p^{i-1}r$$
393
 对上式化简求解 $E[X]$.
34. 为寻找上题的另一种解法，我们令 T_r 表示为获得 r 次连续正面所需要的抛硬币次数.
 (a) 确定 $E[T_r|T_{r-1}]$.
 (b) 写出 $E[T_r]$ 关于 $E[T_{r-1}]$ 的表达式.
 (c) 写出 $E[T_1]$ 的表达式.
 (d) 写出 $E[T_r]$ 的表达式.
35. (a) 证明：
$$E[X]=E[X|X<a]P\{X<a\}+E[X|X\geqslant a]P\{X\geqslant a\}$$
 提示：定义一个适当的随机变量并以其为条件写出 $E[X]$.
 (b) 利用(a)证明马尔可夫不等式，即如果 $P\{X\geqslant 0\}=1$ 成立，则对任意的 $a>0$，有
$$P\{X\geqslant a\}\leqslant\frac{E[X]}{a}$$
36. 一个坛子中装有 a 个白球和 b 个黑球，每次随机地从中取出一个球，直到坛子中只剩下同一颜色的球. 用 $M_{a,b}$ 表示直到上述目标实现时坛子中剩余的球数的期望值，建立 $M_{a,b}$ 的递推关系式，并当 $a=3,b=5$ 时求解.
37. 设坛子中装有 a 个白球和 b 个黑球，随机地取出一球，如果是白球就将其放回，如果是黑球就把它换成白球并放回. 令 M_n 表示经过如上 n 次取球后坛子中白球的期望值.
 (a) 导出递推方程

$$M_{n-1}=\left(1-\frac{1}{a+b}\right)M_n+1$$

(b) 用(a)证明：

$$M_n=a+b-b\left(1-\frac{1}{a+b}\right)^n$$

(c) 第 $n+1$ 次取出的是白球的概率为多少？

38. Y 关于 X_1 和 X_2 的最佳线性预测等于 $a+bX_1+cX_2$，其中 a,b,c 取值使得

$$E[(Y-(a+bX_1+cX_2))^2]$$

达到最小，求 a,b,c

39. Y 关于 X 的最佳二次预测为 $a+bX+cX^2$，其中 a,b,c 取值使得

$$E[(Y-(a+bX+cX^2))^2]$$

达到最小，求 a,b,c

40. 设 X 和 Y 是具有联合正态分布的随机变量，其联合密度函数为

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\times\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}$$

394

(a) 试证在 $X=x$ 的条件下，Y 的条件分布是以 $\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$ 为均值、以 $\sigma_y^2(1-\rho^2)$ 为方差的正态分布.

(b) 证明：$\text{Cov}(X,Y)=\rho$.

(c) 证明 X 和 Y 相互独立的充要条件是 $\rho=0$.

41. 设 X 为一正态随机变量，均值 $\mu=0$，方差 $\sigma^2=1$，令 I 为一独立于 X 的随机变量，且 $P\{I=1\}=\frac{1}{2}=P\{I=0\}$. 定义随机变量 Y 为

$$Y=\begin{cases}X & 如果\ I=1\\ -X & 如果\ I=0\end{cases}$$

也即 Y 等可能地取值 $-X$ 或 X.

(a) X 和 Y 是否独立？

(b) I 和 Y 是否独立？

(c) 证明 Y 服从均值为 0. 方差为 1 的正态分布.

(d) 证明 $\text{Cov}(X,Y)=0$.

42. 由命题 5.1 以及 Y 关于 X 的最佳线性预测为 $\mu_y+\rho\frac{\sigma_y}{\sigma_x}(X-\mu_x)$ 这一事实可知，若 $E[Y|X]=a+bX$，则 $a=\mu_y-\rho\frac{\sigma_y}{\sigma_x}\mu_x$，$b=\rho\frac{\sigma_y}{\sigma_x}$.（为什么？）试直接证明之.

43. 对于随机变量 X 和 Z，证明 $E[(X-Y)^2]=E[X^2]-E[Y^2]$，这里 $Y=E[X|Z]$.

44. 考虑某生物群体，每个个体均可以繁衍后代. 假设每个个体在生命结束时，繁衍出 j 个新后代的概率为 $P_j(j\geqslant 0)$，且每个个体繁衍后代的数目是相互独立的. 最初这一群体的数目记为 X_0，称为第 0 代的生物数目. 所有第 0 代直接繁衍的后代称为第 1 代生物，数目记为 X_1，等等. 用 X_n 表示第 n 代的生物数目. 令 $\mu=\sum_{j=0}^{\infty}jP_j$，$\sigma^2=\sum_{j=0}^{\infty}(j-\mu)^2P_j$ 分别表示一个个体所繁衍的后代数目的均值和方差. 不妨假设 $X_0=1$，即最初只有一个个体.

(a) 证明：$E[X_n]=\mu E[X_{n-1}]$

395

(b) 利用(a)推出 $E[X_n]=\mu^n$.

(c) 证明：$\text{Var}(X_n)=\sigma^2\mu^{n-1}+\mu^2\text{Var}(X_{n-1})$.

(d) 利用(c)推出

$$\operatorname{Var}(X_n)=\begin{cases}\sigma^2\mu^{n-1}\left(\dfrac{\mu^n-1}{\mu-1}\right) & \text{如果 } \mu\neq 1\\ n\sigma^2 & \text{如果 } \mu=1\end{cases}$$

此模型描述的就是著名的分支过程，一个很重要的问题是考虑这个生物群体最终会灭亡的概率问题. 令 π 表示这个群体由单个个体开始，最终会灭亡的概率，即

$$\pi=P\{\text{群体最终会灭亡}\mid X_0=1\}$$

(e) 证明 π 满足

$$\pi=\sum_{j=0}^{\infty}P_j\pi^j$$

提示：以最初的个体所产生的后代为条件进行讨论.

45. 验证表 7-2 中所列的均匀随机变量的矩母函数，并对其微分进一步验证各自的均值和方差.

46. 设 Z 为标准正态随机变量，令 $\mu_n=E[Z^n]$，证明

$$\mu_n=\begin{cases}0 & \text{如果 } n \text{ 为奇数}\\ \dfrac{(2j)!}{2^j j!} & \text{如果 } n=2j\end{cases}$$

提示：对 Z 的矩母函数在 0 点处进行泰勒级数展开可得.

$$E[e^{tZ}]=e^{t^2/2}=\sum_{j=0}^{\infty}\frac{(t^2/2)^j}{j!}$$

47. 设 X 为均值为 μ、方差为 σ^2 的正态随机变量，利用理论练习 46 的结论证明.

$$E[X^n]=\sum_{j=0}^{[n/2]}\frac{\binom{n}{2j}\mu^{n-2j}\sigma^{2j}(2j)!}{2^j j!}$$

上式中$[n/2]$表示 $n/2$ 的整数部分，并在 $n=1$ 和 $n=2$ 时验证所得结论.

48. 如果 $Y=aX+b$，其中 a、b 为常数，试用 X 的矩母函数来表示 Y 的矩母函数.

49. 我们称正随机变量 X 服从参数为 μ 和 σ^2 的对数正态分布，如果 $\log(X)$服从均值为 μ、方差为 σ^2 的正态分布. 试利用正态分布的矩母函数求对数正态分布的矩母函数.

396

50. 设随机变量 X 的矩母函数为 $M(t)$，定义 $\Psi(t)=\log M(t)$，试证明

$$\Psi''(t)|_{t=0}=\operatorname{Var}(X)$$

51. 设 $X_1,X_2,\cdots,X_n$ 为相互独立的均值为$\frac{1}{\lambda}$的指数分布随机变量，试利用表 7-2 求 $\sum_{i=1}^{n}X_i$ 的概率分布.

52. 证明如何利用随机变量 X 和 Y 的矩母函数来确定 $\operatorname{Cov}(X,Y)$.

53. 设 $X_1,X_2,\cdots,X_n$ 服从多元正态分布，证明 $X_1,X_2,\cdots,X_n$ 为相互独立随机变量的充要条件是 $\operatorname{Cov}(X_i,X_j)=0$，$i\neq j$.

54. 如果随机变量 Z 服从标准正态分布，求 $\operatorname{Cov}(Z,Z^2)$.

自测题与练习

1. 考虑一张有 m 个名字的表，其中相同的名字出现不止一次. 用 $n(i)$表示处于第 i 个位置的名字出现的次数，其中 $i=1,2,\cdots,m$. 假设共有 d 个不同的名字.

(a) 写出 d 关于 m、$n(i)(i=1,2,\cdots,m)$的表达式.

进一步假设 U 服从$(0,1)$上的均匀分布，并令 $X=[mU]+1$.

(b) 求 X 的概率质量函数.

(c) 证明 $E[m/n(X)]=d$.

2. 一个坛子中装有 n 个白球和 m 个黑球，每次随机地从中取出一个球，试求一个白球紧接着一个黑球后被取出的这种情况发生的期望数.
3. 有 10 对夫妇，坐在 5 张不同的桌子旁，每 4 个人一张桌子 .
   (a) 假如随机地选择所坐的位置，求坐在同一张桌子旁的夫妇数的期望值.
   (b) 假如随机地选择 2 个男人和 2 个女人坐在一张桌子旁，求坐在同一张桌子旁的夫妇数的期望值.
4. 掷一个骰子直到所有的点数都至少出现一次，求结束时 1 点出现次数的期望值.
5. 一套卡片共有 $2n$ 张，其中有 n 张红卡片和 n 张黑卡片，将这些卡片充分混合后每次翻开一张. 若到某次已经翻开的红卡片不比黑卡片少，则每一张红卡片，我们赢 1 美元. (例如，当 $n=2$ 时，如果翻开结果为 r,b,r,b，则我们就赢 2 美元.)求我们赢的总钱数的期望值.
6. 设 $A_1,\cdots,A_n$ 表示 n 个随机事件，N 表示事件发生的个数，另外，令 $I=1$ 表示这些事件发生，否则令 $I=0$，证明伯努利不等式，即

$$P(A_1\cdots A_n)\geqslant\sum_{i=1}^{n}P(A_i)-(n-1)$$

   提示：首先讨论 $N\leqslant n-1+I$.

397

7. 从数集 $1,2,\cdots n$,中随机取出 k 个数，用 X 表示所取数中的最小值，试利用例 2 m 中的方法求 $E[X]$.
8. 一架到站的飞机上共有 r 名乘客，其中随行携带 j 件行李的乘客有 n_j 名，$\sum_j n_j=r$. 假设飞机降落后所有行李运出的先后顺序是随机的，且假定乘客拿到自己的行李后就马上离开. 设乘客 Sanchez 托运有 j 件行李，试求在他之后离开的乘客数的期望值.
*9. 有 19 颗钉子随机地钉在一个半径为 1 的圆盘的边缘上，证明无论钉子的排列位置如何，总可以找到一段长为 1 的弧上至少有 4 颗钉子.
10. 设 X 服从均值为 λ 的泊松分布，试证明在 λ 不是太小的情况下，有

$$\mathrm{Var}(\sqrt{X})\approx 0.25$$

   提示：利用理论练习 4 的结论来近似 $E[\sqrt{X}]$.
11. 假设在自测题与练习 3 中 10 对夫妇坐在 7 张桌子旁，其中 3 张桌子有 4 个座位，4 张桌子有 2 个座位. 如果所有人被随机地安排座位，试求坐在同一张桌子旁的夫妇数的数学期望.
12. 编号从 1 到 n 的 n 个人按如下的规则进入一家公司工作：首先 1 号进入公司而后作为招聘者招聘 2 号；当 2 号进入公司后，1 号和 2 号等可能地被选中作为招聘者去招聘 3 号；当 3 号进入公司后，1 号、2 号和 3 号等可能地被选中作为招聘者去招聘 4 号；等等，直到所有 n 个人都进入公司.
    (a) 试求在 $1,2,\cdots,n$ 中从没有被选中作为招聘者的人数的期望值.
    (b) 当 $n=5$ 时，试导出在 $1,2,\cdots,n$ 中从没有被选中作为招聘者的人数的方差的表达式.
13. 一支由 9 人组成的篮球队，其中包括 2 名中锋、3 名前锋和 4 名后卫. 如果这些队员被随机地按每 3 人一组分为三组.
    (a) 求所分成的三组中包含所有三种类型的球员的小组数的期望值.
    (b) 求所分成的三组中包含所有三种类型的球员的小组数的方差.
14. 一副扑克牌共有 52 张，充分混合后从中随机取出 13 张，用 X 和 Y 分别表示其中的 A 的数目和黑桃的数目.
    (a) 证明 X 和 Y 是不相关的.
    (b) 它们是独立的吗?
15. 一个箱子中的所有硬币都具有一定的面值. 假设取出并抛一枚面值为 p 的硬币正面朝上的概率为 p，当随机地从箱子中取出一枚硬币时，其面值服从(0,1)上的均匀分布. 假设在取出硬币抛之前都要预测它是正面朝上还是反面朝上，预测准确将赢得 1 美元，错误将输 1 美元.

(a) 假如你不知道硬币面值，试求你赢得钱数的期望值.

(b) 假设在每次抛硬币之前都知道其面值 p，你应做出怎样的预测?

(c) 在(b)的情况下，试求所赢得钱数的期望值.

16. 在自测题 1 中我们给出了如何利用(0,1)上的均匀分布随机变量的值(通常称为随机数)来获得一个随机变量的值，使得其均值等于所列的名字中不同名字的数目. 然而它需要随机选中一个位置，然后确定这个位置的名字出现的次数. 下面介绍在有名字重复出现次数较多时的一种更有效的方法. 和前面一样，首先选中如题 1 中的随机变量 X，然后从头开始找与位置 X 的名字相同的名字，如果在位置 X 之前就找到了记 $I=0$，如果位置 X 之前找不到记 $I=1$，试证明 $E[mI]=d$.

398

提示：可以用条件数学期望求 $E[I]$.

17. m 条指令被随机存放在 n 个不同的单元里，且每条指令被存放在第 j 个单元里的概率为 p_j，$j=1,2,\cdots,n$. 如果一条指令存放在某个单元时这个单元已经被其他的指令占用了，我们就称为一个冲突发生，试求冲突发生次数的期望值 .

18. 有 n 个 1 和 m 个 0 随机地排列在一起，所有相同的数连续排列我们就称为一个游程. 用 L 表示第一个游程中数字的长度，即如果前 k 个数都一样(都为 1 或都为 0)，则 $L\geqslant k$，试求 $E[L]$.

19. 在标号为 H 的箱子里有 n 个球，标号为 T 的箱子里有 m 个球. 现抛一枚硬币，其正面朝上的概率为 p，反面朝上的概率为 $1-p$. 每次如果抛硬币的结果是正面朝上，则从标号为 H 的箱子中取出一个球；反面朝上就从标号为 T 的箱子中取出一个球. 如果一个箱子被取空就不再取球而继续下一次抛硬币，试求直到两个箱子均被取空时需要抛硬币次数的期望值.

提示：以前 $n+m$ 次抛硬币正面朝上的次数为条件.

20. 设 X 是分布函数为 F 的非负随机变量，记 $\bar{F}(x)=1-F(x)$，试证明

$$E[X^n]=\int_0^{\infty} x^{n-1}\bar{F}(x)\mathrm{d}x$$

提示：利用 $X^n=n\int_0^X x^{n-1}\mathrm{d}x=n\int_0^{\infty} x^{n-1}I_X(x)\mathrm{d}x$，这里

$$I_X(x)=\begin{cases}1 & \text{如果 } x<X\\ 0 & \text{否则}\end{cases}$$

399

第8章 极限定理

8.1 引言

概率论中最重要的理论结果是极限定理. 在这些定理中，尤为重要的是大数定律(law of large number)或中心极限定理(central limit theorem). 通常，把在所声明的条件下，一个随机变量序列的算术平均值(按某种意义)收敛于所希望的平均值的定理，归为大数定律. 另一方面，中心极限定理则是关于确定在什么条件下，大量的随机变量之和具有近似正态的概率分布.

8.2 切比雪夫不等式与弱大数定律

本节我们先证明一个称为马尔可夫(Markov)不等式的结果.

命题 2.1(马尔可夫不等式) 设 X 为只取非负值的随机变量，则对任意实数 $a>0$，有

$$P\{X \geqslant a\} \leqslant \frac{E[X]}{a}$$

证明 对 $a>0$，令

$$I = \begin{cases} 1 & \text{如果 } X \geqslant a \\ 0 & \text{其他} \end{cases}$$

400

又因为 $X \geqslant 0$，所以

$$I \leqslant \frac{X}{a}$$

对上式两端取数学期望得

$$E[I] \leqslant \frac{E[X]}{a}$$

因为 $E[I]=P\{X \geqslant a\}$，从而结论成立. ■

命题 2.2 是命题 2.1 的推论.

命题 2.2(切比雪夫不等式) 设 X 为随机变量，具有有限均值 μ 和方差 σ^2，则对任意值 $k>0$，

$$P\{|X-\mu| \geqslant k\} \leqslant \frac{\sigma^2}{k^2}$$

证明 由于 $(X-\mu)^2$ 是非负随机变量，故可用马尔可夫不等式(令 $a=k^2$)得

$$P\{(X-\mu)^2 \geqslant k^2\} \leqslant \frac{E[(X-\mu)^2]}{k^2} \tag{2-1}$$

但是，由于 $(X-\mu)^2 \geqslant k^2$ 的充要条件为 $|X-\mu| \geqslant k$，因此式(2-1)等价于

$$P\{|X-\mu| \geqslant k\} \leqslant \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

命题证毕. ■

马尔可夫不等式和切比雪夫不等式的重要性在于，当我们仅仅知道概率分布的均值，或者同时知道其均值及方差时，可以利用它们得到概率的界. 当然，如果已知随机变量实际的分布，那么所求的概率就可以确切地算出来，我们也就没有必要求助于它的界了.

例 2a 设已知某工厂一周的产量是均值为 50 的随机变量.

(a) 关于这一周的产量将超过 75 的概率，你有什么看法？

(b) 若已知周产量的方差为 25，则关于这一周的产量将在 40 到 60 之间的概率，你有什么看法？

401

解 设 X 表示周产量.

(a) 由马尔可夫不等式得

$$P\{X>75\}\leqslant\frac{E[X]}{75}=\frac{50}{75}=\frac{2}{3}$$

(b) 由切比雪夫不等式得

$$P\{|X-50|\geqslant 10\}\leqslant\frac{\sigma^2}{10^2}=\frac{1}{4}$$

因此

$$P\{|X-50|<10\}\geqslant 1-\frac{1}{4}=\frac{3}{4}$$

可见这一周的产量在 40 到 60 之间的概率至少为 0.75. ■

由于切比雪夫不等式对任何分布的随机变量 X 都是成立的，因此，在多数情况下，我们不能指望所得的概率的界非常接近于真正的概率. 例如，考虑例 2b.

例 2b 如果 X 服从(0，10)上的均匀分布，则当 $E[X]=5$，$\mathrm{Var}(X)=\frac{25}{3}$时，由切比雪夫不等式知

$$P\{|X-5|>4\}\leqslant\frac{25}{3(16)}\approx 0.52$$

但确切的结果为

$$P\{|X-5|>4\}=0.20$$

可见，虽然切比雪夫不等式是正确的，但它给出的上界并不是特别接近于实际的概率.

类似地，如果 X 是均值为 μ、方差为 σ^2 的正态随机变量，则由切比雪夫不等式得

$$P\{|X-\mu|>2\sigma\}\leqslant\frac{1}{4}$$

但实际的概率等于

$$P\{|X-\mu|>2\sigma\}=P\left\{\left|\frac{X-\mu}{\sigma}\right|>2\right\}=2[1-\Phi(2)]\approx 0.045\ 6$$ ■

切比雪夫不等式经常作为证明结论的理论工具. 我们先用命题 2.3，然后用更为重要的弱大数定律来说明这一点.

命题 2.3 若 $\mathrm{Var}(X)=0$，则

$$P\{X=E[X]\}=1$$

402

换句话说，方差等于 0 的随机变量与以概率 1 等于常数的随机变量是一样的.

证明 根据切比雪夫不等式，对任意 $n\geqslant 1$，我们有

$$P\left\{|X-\mu|>\frac{1}{n}\right\}=0$$

令 $n\to\infty$并用概率的连续性即得

$$0=\lim_{n\to\infty}P\left\{|X-\mu|>\frac{1}{n}\right\}=P\left\{\lim_{n\to\infty}\left\{|X-\mu|>\frac{1}{n}\right\}\right\}=P\{X\neq\mu\}$$

命题得证. ■

定理 2.1(弱大数定律) 设 $X_1, X_2\cdots$是独立同分布的随机变量序列，且都具有有限的数学期望 $E[X_i]=\mu$，则对任意 $\varepsilon>0$，当 $n\to\infty$时，有

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|\geqslant\varepsilon\right\}\to 0$$

证明 我们将只在随机变量的方差 σ^2 有限这一附加条件下证明上述结果. 因为

$$E\left[\frac{X_1+\cdots+X_n}{n}\right]=\mu, \mathrm{Var}\left(\frac{X_1+\cdots+X_n}{n}\right)=\frac{\sigma^2}{n}$$

故由切比雪夫不等式得

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|\geqslant\varepsilon\right\}\leqslant\frac{\sigma^2}{n\varepsilon^2}$$

定理得证. ■

弱大数定律的特殊情况，即 X_i 是 0－1(即伯努利)随机变量，最初被詹姆士·伯努利(James Bernoulli)所证明，有关论述及定理的证明发表在他的著作《Ars Conjectandi》中，这本书是在他逝世之后 8 年(1713 年)由其侄子尼古拉斯·伯努利(Nicholas Bernoulli)出版的. 应当指出，詹姆士·伯努利在世时，切比雪夫不等式尚不知道，他是凭借十分巧妙的证明方法得到其结果的. 定理 2.1 所述的弱大数定律的一般形式，是由俄国数学家辛钦(Khintchine)证明的.

8.3 中心极限定理

中心极限定理是概率论中最著名的结果之一. 粗略地说，中心极限定理指出，大量的独立随机变量之和具有近似于正态的分布. 因此，它不仅提供了计算独立随机变量之和的近似概率的简单方法，而且有助于解释为什么很多自然群体的经验频率呈现出钟形(即正态)曲线这一值得注意的事实. 403

中心极限定理的最简单形式如下：

定理 3.1(中心极限定理) 设 $X_1, X_2, \cdots$是独立同分布的随机变量序列，各有均值 μ 及方差 σ^2，则当 $n\to\infty$时，

$$\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$$

的分布趋于标准正态分布. 也就是说，当 $n\to\infty$时，对于$-\infty<a<\infty$，有

$$P\left\{\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\leqslant a\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a}\mathrm{e}^{-x^2/2}\mathrm{d}x$$

证明中心极限定理的关键是下面的引理，我们不加证明地陈述如下：

引理 3.1 设 $Z_1, Z_2, \cdots$为随机变量序列，其分布函数为 F_{Z_n}，矩母函数为 M_{Z_n}，$n\geqslant 1$ 又设 Z 为随机变量，其分布函数为 F_Z，矩母函数为 M_Z. 如果对一切 t，$M_{Z_n}(t)\to M_Z(t)$，则对一切使 $F_Z(t)$在 t 点连续的 t，有 $F_{Z_n}(t)\to F_Z(t)$.

如果设 Z 为标准正态随机变量，则由于 $M_Z(t)=\mathrm{e}^{t^2/2}$，故由引理 3.1 得知，若 $n\to\infty$时 $M_{Z_n}(t)\to\mathrm{e}^{t^2/2}$，则当 $n\to\infty$时 $F_{Z_n}(t)\to\Phi(t)$.

至此我们已为证明中心极限定理做好了准备.

中心极限定理的证明 首先设$\mu=0$, $\sigma^2=1$. 我们只对X_i的矩母函数$M(t)$存在且有限的情形证明上述定理. $X_i/\sqrt{n}$的矩母函数为

404

$$E\left[\exp\left\{\frac{tX_i}{\sqrt{n}}\right\}\right]=M\left(\frac{t}{\sqrt{n}}\right)$$

从而$\sum_{i=1}^{n}X_i/\sqrt{n}$的矩母函数为$\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$. 令

$$L(t)=\log M(t)$$

又

$$L(0)=0$$

$$L'(0)=\frac{M'(0)}{M(0)}=\mu=0$$

$$L''(0)=\frac{M(0)M''(0)-[M'(0)]^2}{[M(0)]^2}=E[X^2]=1$$

为证明定理成立，首先必须证明当$n\to\infty$时，$[M(t/\sqrt{n})]^n\to e^{t^2/2}$，或者等价地，证明当$n\to\infty$时，$nL(t/\sqrt{n})\to t^2/2$. 为此，注意

$$\begin{aligned}\lim_{n\to\infty}\frac{L(t/\sqrt{n})}{n^{-1}}&=\lim_{n\to\infty}\frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} \quad \text{由洛必达(L'Hospital)法则}\\&=\lim_{n\to\infty}\left[\frac{L'(t/\sqrt{n})t}{2n^{-1/2}}\right]\\&=\lim_{n\to\infty}\left[\frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}}\right] \quad \text{又由洛必达法则}\\&=\lim_{n\to\infty}\left[L''(t/\sqrt{n})\frac{t^2}{2}\right]=\frac{t^2}{2}\end{aligned}$$

从而当$\mu=0$, $\sigma^2=1$时，中心极限定理成立. 对一般的情况，考虑标准化随机变量$X_i^*=(X_i-\mu)/\sigma$, 因$E[X_i^*]=0$, $\mathrm{Var}(X_i^*)=1$, 对X_i^*应用上面的结果，定理即得证. ■

注 虽然定理3.1仅指出，对任意a,

$$P\left\{\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\leqslant a\right\}\to\Phi(a)$$

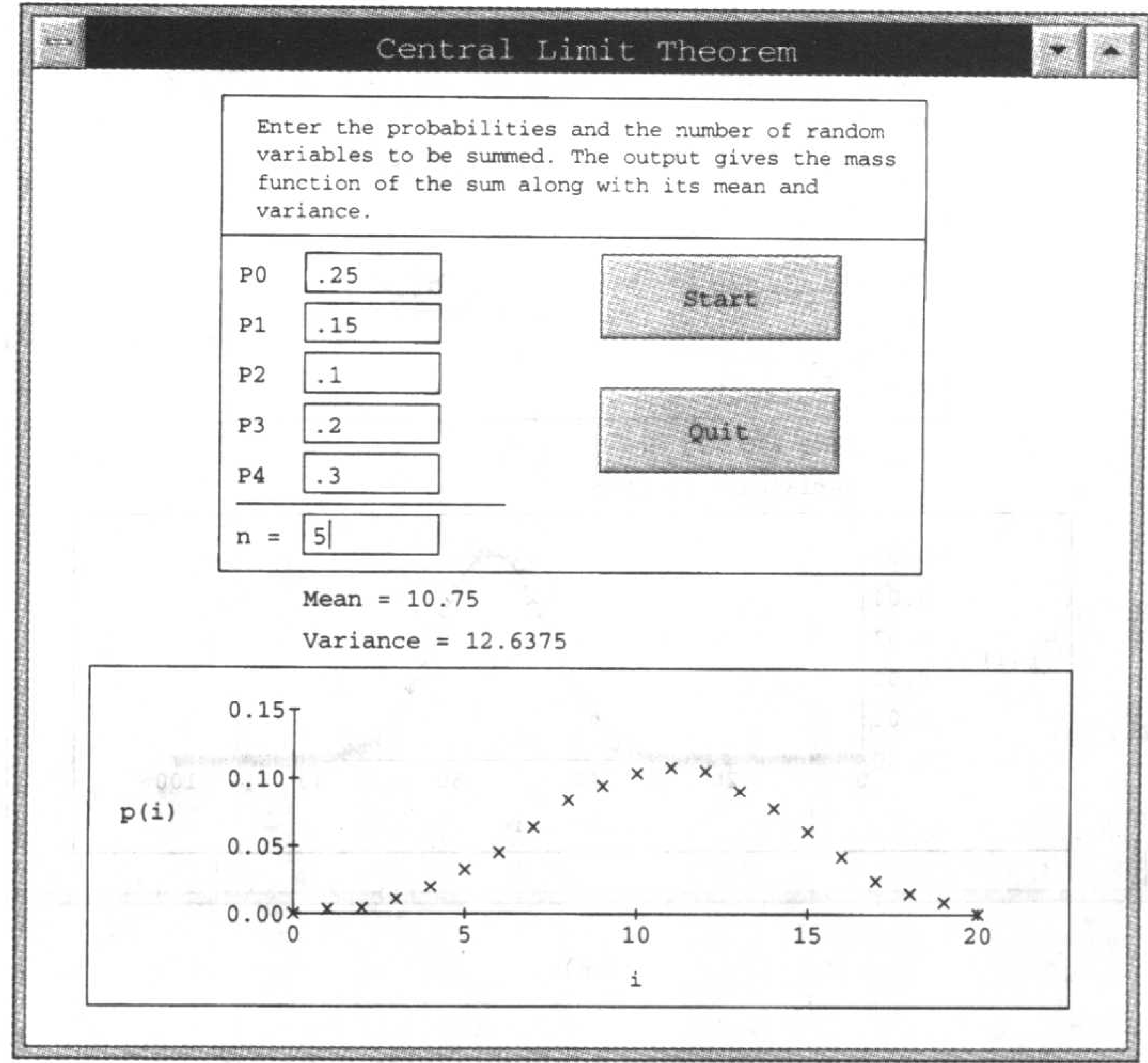
实际上可以证明，上述收敛对a是一致的.[我们称$f_n(a)\to f(a)$是一致的，如果对任意$\varepsilon>0$，存在N，使得当$n\geqslant N$时，对所有的a都有$|f_n(a)-f(a)|<\varepsilon$.]

405

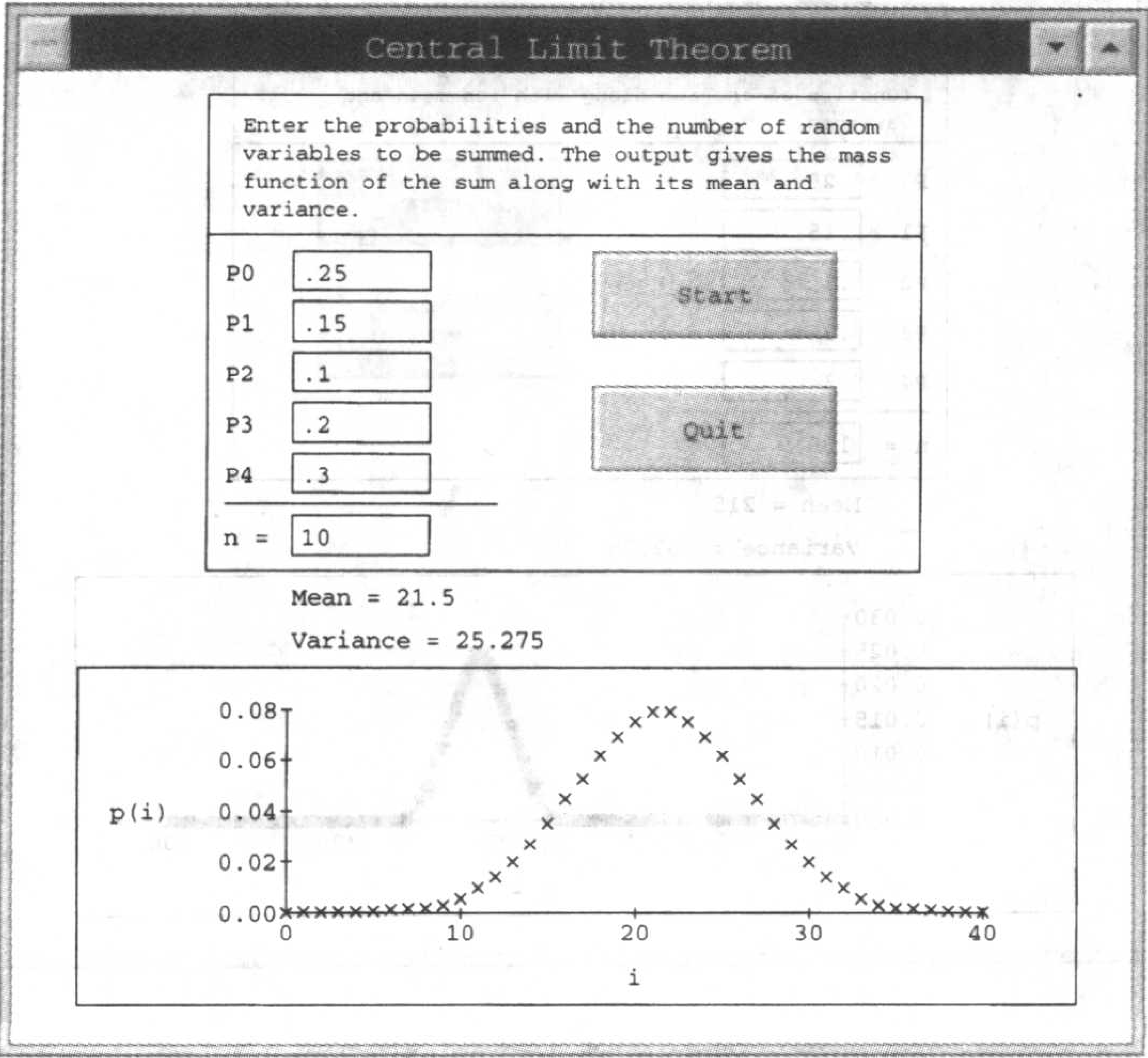
大约在1733年，棣莫弗对X_i为贝努利随机变量且$p=\frac{1}{2}$的特殊情况证明了中心极限定理，这是中心极限定理最早的陈述. 后来拉普拉斯把它推广到p是任意的场合.(因为二项随机变量可视为n个独立同分布的伯努利随机变量之和，所以这就证明了5.4节中所指出的关于二项随机变量的正态近似性质.)拉普拉斯还发现了中心极限定理的更一般形式，即定理3.1. 但其证明不太严格，实际上也很难严格. 中心极限定理真正严格的证明，是俄国数学家李雅普诺夫(Liapounoff)在1901～1902年间首先给出的.

在网站上，这个重要的定理被解释为中心极限定理模块. 网站中绘出了取值为0,1,2,3,4

的 n 个独立同分布随机变量之和的密度函数. 要使用它，只需键入概率质量函数和 n 的期望值. 当(a)$n=5$，(b)$n=10$，(c)$n=25$，(d)$n=100$ 时，图 8-1 给出了特定概率质量函数的图表.

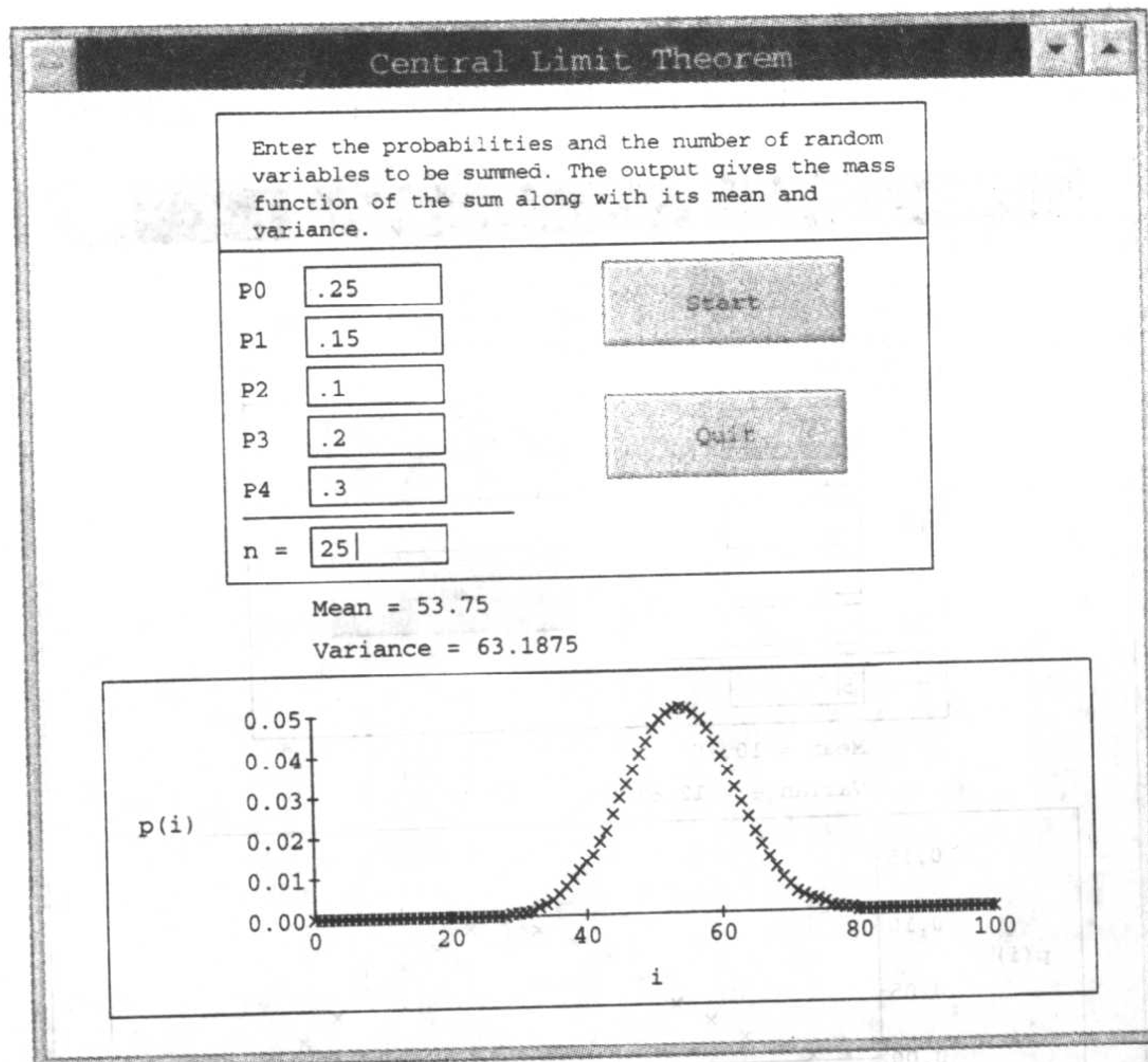


a)



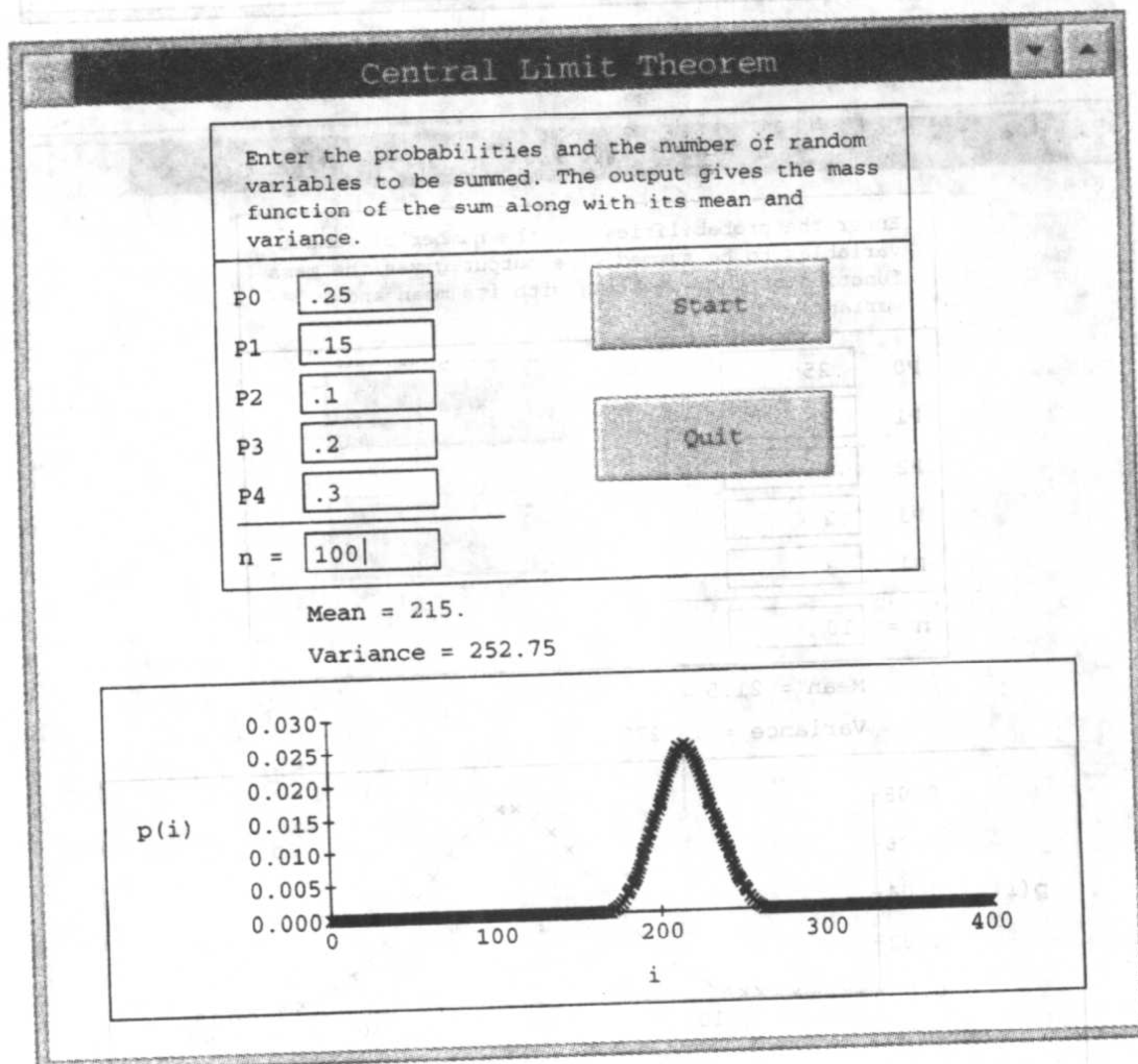
b)

图 8-1



Central Limit Theorem
Enter the probabilities and the number of random variables to be summed. The output gives the mass function of the sum along with its mean and variance.
P0 .25
P1 .15
P2 .1
P3 .2
P4 .3
n = 25
Start
Quit
Mean = 53.75
Variance = 63.1875
p(i)
0.05
0.04
0.03
0.02
0.01
0.00
0
20
40
60
80
100
i

c)



Central Limit Theorem
Enter the probabilities and the number of random variables to be summed. The output gives the mass function of the sum along with its mean and variance.
P0 .25
P1 .15
P2 .1
P3 .2
P4 .3
n = 100
Start
Quit
Mean = 215.
Variance = 252.75
p(i)
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
100
200
300
400
i

d)

图 8-1 （续）

例 3a 某天文学家致力于测量从他的观测台到某一个星的距离(单位：光年). 虽然这位天文学家拥有测量技术，但他知道，由于大气条件的变化以及正常的测量误差，使每次测量都不能测得正确的距离，只能得到其近似估计值. 为了得到结果，他打算进行一系列测量，然后取这些测量值的平均值作为实际距离的估计值. 如果这位天文学家确信，上述一系列测量值是独立同分布的随机变量，有共同的均值 d(即实际距离)及同一方差 4(单位：光年). 试问他必须进行多少次测量，才能有理由确信其距离估计值的误差在 ± 0.5 光年以内?

解 设天文学家决定进行 n 次测量. 令 $X_1, X_2, \cdots, X_n$ 表示这 n 次测量的值，则由中心极限定理可知

$$Z_n = \frac{\sum_{i=1}^{n} X_i - nd}{2\sqrt{n}}$$

近似地具有标准正态分布，因此

$$P\left\{-0.5 \leqslant \frac{\sum_{i=1}^{n} X_i}{n} - d \leqslant 0.5\right\} = P\left\{-0.5\frac{\sqrt{n}}{2} \leqslant Z_n \leqslant 0.5\frac{\sqrt{n}}{2}\right\}$$

$$\approx \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

于是，如果他希望，比如说，以 95%的把握确信其估计值精确到 0.5 光年以内，他就要进行 n^* 次测量，其中 n^* 满足

$$2\Phi\left(\frac{\sqrt{n^*}}{4}\right) - 1 = 0.95 \text{ 或者 } \quad \Phi\left(\frac{\sqrt{n^*}}{4}\right) = 0.975$$

于是由第 5 章表 5-1 知

$$\frac{\sqrt{n^*}}{4} = 1.96 \text{ 或者 } n^* = (7.84)^2 \approx 61.47$$

因为 n^* 不是整数，故他必须做 62 次观测.

但必须指出，上面的分析是在这样的条件下进行的：当 $n=62$ 时正态近似已是好的近似. 尽管这通常没有什么问题，但一般说来，“好”的近似需要多大的 n 的问题是与 X_i 的分布有关的. 如果天文学家注意到这一点，又不想凭侥幸做事，他仍然可以借助切比雪夫不等式解决问题. 因为

406
~
409

$$E\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = d \qquad \mathrm{Var}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{4}{n}$$

由切比雪夫不等式得

$$P\left\{\left|\sum_{i=1}^{n} \frac{X_i}{n} - d\right| > 0.5\right\} \leqslant \frac{4}{n(0.5)^2} = \frac{16}{n}$$

因此，他如果做 $n = 16/0.05 = 320$ 次观测，就能以 95%的把握确信其估计值精确到 0.5 光年以内. ■

例 3b 报名听心理学课的学生人数是均值为 100 的泊松随机变量. 负责这门课程的教授决定，如果报名的人数不少于 120，就分成两个班讲授. 而如果少于 120 人，就集中在一个班讲授. 试问该教授将讲授两个班的概率是多少?

解 精确的解 $e^{-100}\sum_{i=120}^{\infty}(100)^i/i!$没有给出具体的数字答案. 但如想到均值为 100 的泊松随机变量等于 100 个均值为 1 的独立泊松随机变量之和，我们就可以利用中心极限定理求其近似解. 设 X 表示报名听课的学生人数，我们有

$$P\{X \geqslant 120\} = P\{X \geqslant 119.5\} \quad （连续性修正）$$
$$= P\left\{\frac{X-100}{\sqrt{100}} \geqslant \frac{119.5-100}{\sqrt{100}}\right\}$$
$$\approx 1-\Phi(1.95) \approx 0.0256$$

其中我们用到了泊松随机变量的方差等于其均值这一事实. ■

例 3c 掷 10 个均匀骰子，求掷出的点数之和在 30 到 40 之间的近似概率.

解 设 X_i 表示第 i 个骰子掷出的点数，其中 $i=1, 2, \cdots, 10$. 因为 $E[X_i]=\frac{7}{2}$，$\mathrm{Var}(X_i)$

410

$=E[X_i^2]-(E[X_i])^2=\frac{35}{12}$，由中心极限定理得

$$P\{29.5 \leqslant X \leqslant 40.5\} = P\left\{\frac{29.5-35}{\sqrt{350/12}} \leqslant \frac{X-35}{\sqrt{350/12}} \leqslant \frac{40.5-35}{\sqrt{350/12}}\right\}$$
$$\approx 2\Phi(1.0184)-1 \approx 0.692$$ ■

例 3d 设 $X_i(i=1, 2, \cdots, 10)$ 为独立随机变量，且都服从$(0, 1)$ 上的均匀分布，计算 $P\left\{\sum_{i=1}^{10} X_i > 6\right\}$的近似值.

解 因为 $E[X_i]=1/2$，$\mathrm{Var}(X_i)=1/12$，由中心极限定理知

$$P\left\{\sum_{1}^{10} X_i > 6\right\} = P\left\{\frac{\sum_{1}^{10} X_i - 5}{\sqrt{10(1/12)}} > \frac{6-5}{\sqrt{10(1/12)}}\right\}$$
$$\approx 1-\Phi(\sqrt{1.2}) \approx 0.1367$$

可见，出现 $\sum_{i=1}^{10} X_i$ 大于 6 的次数只占 14%. ■

当随机变量 X_i 独立但不一定同分布时，中心极限定理也成立. 其中之一但不是最一般的形式如下：

定理 3.2（独立随机变量的中心极限定理） 设 $X_1, X_2, \cdots$ 为随机变量序列，其均值及方差分别为 $\mu_i=E[X_i]$，$\sigma_i^2=\mathrm{Var}(X_i)$. 如果(a) X_i 一致有界，也就是说，若对某 M 有 $P\{|X_i|<M\}=1$ 对一切 i 成立；(b) $\sum_{i=1}^{\infty}\sigma_i^2=\infty$，则当 $n \to \infty$ 时，

$$P\left\{\frac{\sum_{i=1}^{n}(X_i-\mu_i)}{\sqrt{\sum_{i=1}^{n}\sigma_i^2}} \leqslant a\right\} \to \Phi(a)$$

411

历 史 注 记

拉普拉斯(Pierre Simon, Marquis de Laplace)

法国数学家拉普拉斯最先提出并证明中心极限定理,他通过对测量误差的观测(通常被认为是大量细小的力的合力)得出其趋向于正态分布的结论. 拉普拉斯也是一个著名的天文学家(并被称为"法国的牛顿"),他是早期概率统计研究中最伟大的贡献者之一. 拉普拉斯还是在日常生活中使用概率的推广者. 他坚信概率论的重要性,正如下列从他的著作《Analytical Theory of Probability》中引证的那样:

> "实际上,我们知道的常识只是概率论归纳为计算的框架;通常在不能计算的时候,直觉上它使我们有理由相信它的精确度……这门不平常的科学发起了对机会对策的研究,它应该成为人类知识中最重要的学科……对大多数人来说,生命中最重要的问题其实只是概率的问题."

中心极限定理的应用表明测量误差近似于正态分布,这对科学做出了非常重要的贡献. 实际上,在 17 世纪和 18 世纪,中心极限定理通常被称为误差频率定律. 我们来看看弗朗西斯·高尔顿(Francis Galton)是如何说的(引自他 1889 年发表的著作《Natural Inheritance》):

> "我几乎一无所知,从而更倾向于认为'误差频率定律'表述宇宙规则具有非常完美的形式. 如果希腊人知道此定律的话,他们一定会将其个性化与神化. 在极度混乱中,它始终以平静和谦逊的姿态支配着这个宇宙. 民众越多,无政府状态越严重,它的影响就会越完美. 它是最完美的无理性定律."

8.4 强大数定律

强大数定律(strong law of large number)也许是概率论中最著名的结果. 它说明了独立同分布随机变量序列的算术平均值以概率 1 收敛于该分布的均值. [412]

定理 4.1(强大数定律) *设 $X_1, X_2, \cdots$ 为独立同分布的随机变量序列,具有有限的均值 $\mu=E[X_i]$,则当 $n\to\infty$ 时,以概率 1 有*

$$\frac{X_1+X_2+\cdots+X_n}{n}\to\mu^{\ominus}$$

作为强大数定律的一个应用,设进行一系列独立重复试验,令 E 表示试验的某一固定事件,记 $P(E)$ 为 E 在各次试验中出现的概率. 又令

$$X_i=\begin{cases}1 & \text{如果 } E \text{ 出现在第 } i \text{ 次试验}\\ 0 & \text{如果 } E \text{ 不出现在第 } i \text{ 次试验}\end{cases}$$

由强大数定律知,以概率 1 有

$$\frac{X_1+X_2+\cdots+X_n}{n}\to E(X)=P(E) \tag{4-1}$$

⊖ 即强大数定律表示

$$P\{\lim_{n\to\infty}(X_1+\cdots+X_n)/n=\mu\}=1$$

因为 $X_1+\cdots+X_n$ 表示在前 n 次试验中，事件 E 出现的次数，故式(4-1)可解释为事件 E 发生的次数所占的比例的极限，以概率 1 恰为 $P(E)$.

尽管不用任何假设就可以证明此定理，但我们证明强大数定律仍然要假设随机变量 X_i 的四阶矩有限. 也就是说，假设 $E[X_i^4]=K<\infty$.

强大数定律的证明 首先假设 X_i 的均值 μ 等于 0. 令 $S_n=\sum_{i=1}^{n}X_i$，考虑

$$E(S_n^4)=E[(X_1+\cdots+X_n)(X_1+\cdots+X_n)\times(X_1+\cdots+X_n)(X_1+\cdots+X_n)]$$

将上式右端展开，所得结果中有以下各项：

$$X_i^4,\ X_i^3X_j,\ X_i^2X_j^2,\ X_i^2X_jX_k,\ X_iX_jX_kX_l$$

413

其中 i, j, k, l 互不相同. 因为所有 X_i 的均值 μ 都等于 0，从而由独立性知

$$E[X_i^3X_j]=E[X_i^3]E[X_j]=0$$
$$E[X_i^2X_jX_k]=E[X_i^2]E[X_j]E[X_k]=0$$
$$E[X_iX_jX_kX_l]=0$$

如果给定一对 i, j，则展开式中就有 $\binom{4}{2}=6$ 项 $X_i^2X_j^2$. 因此分别对上述展开的乘积的每一项取均值有

$$E[S_n^4]=nE[X_i^4]+6\binom{n}{2}E[X_i^2X_j^2]=nK+3n(n-1)E[X_i^2]E[X_j^2]$$

其中再次用到了独立性假设. 又因为

$$0\leqslant \mathrm{Var}(X_i^2)=E[X_i^4]-(E[X_i^2])^2$$

由此可知

$$(E[X_i^2])^2\leqslant E[X_i^4]=K$$

因此，由以上结果可知

$$E[S_n^4]\leqslant nK+3n(n-1)K$$

即

$$E\left[\frac{S_n^4}{n^4}\right]\leqslant\frac{K}{n^3}+\frac{3K}{n^2}$$

进而可得

$$E\left[\sum_{n=1}^{\infty}\frac{S_n^4}{n^4}\right]=\sum_{n=1}^{\infty}E\left[\frac{S_n^4}{n^4}\right]<\infty$$

从上式可得，以概率 1 有 $\sum_{n=1}^{\infty}S_n^4/n^4<\infty$.（因为如果此和式无限的概率为正，那么其均值无限.）由级数收敛可得其第 n 项趋于 0；因此以概率 1 有

$$\lim_{n\to\infty}\frac{S_n^4}{n^4}=0$$

但若 $S_n^4/n^4=(S_n/n)^4$ 趋于 0，则必有 S_n/n 趋于 0. 从而当 $n\to\infty$时，以概率 1 有

$$\frac{S_n}{n}\to 0$$

414

当 X_i 的均值 μ 不等于 0 时，对 $X_i-\mu$ 重复上述讨论可得，以概率 1 有

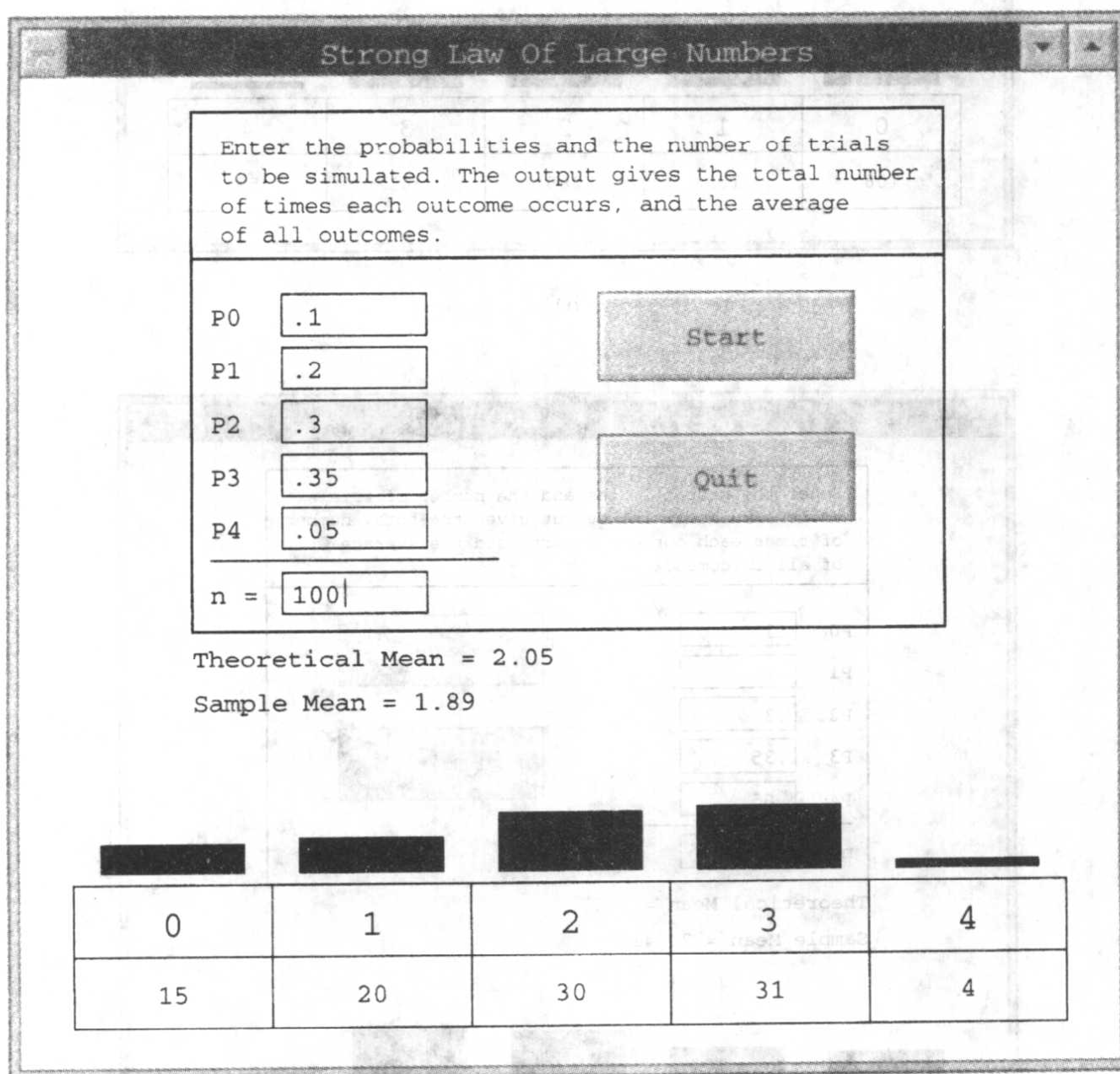
$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{(X_i-\mu)}{n}=0$$

或者等价地，

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{X_i}{n}=\mu$$

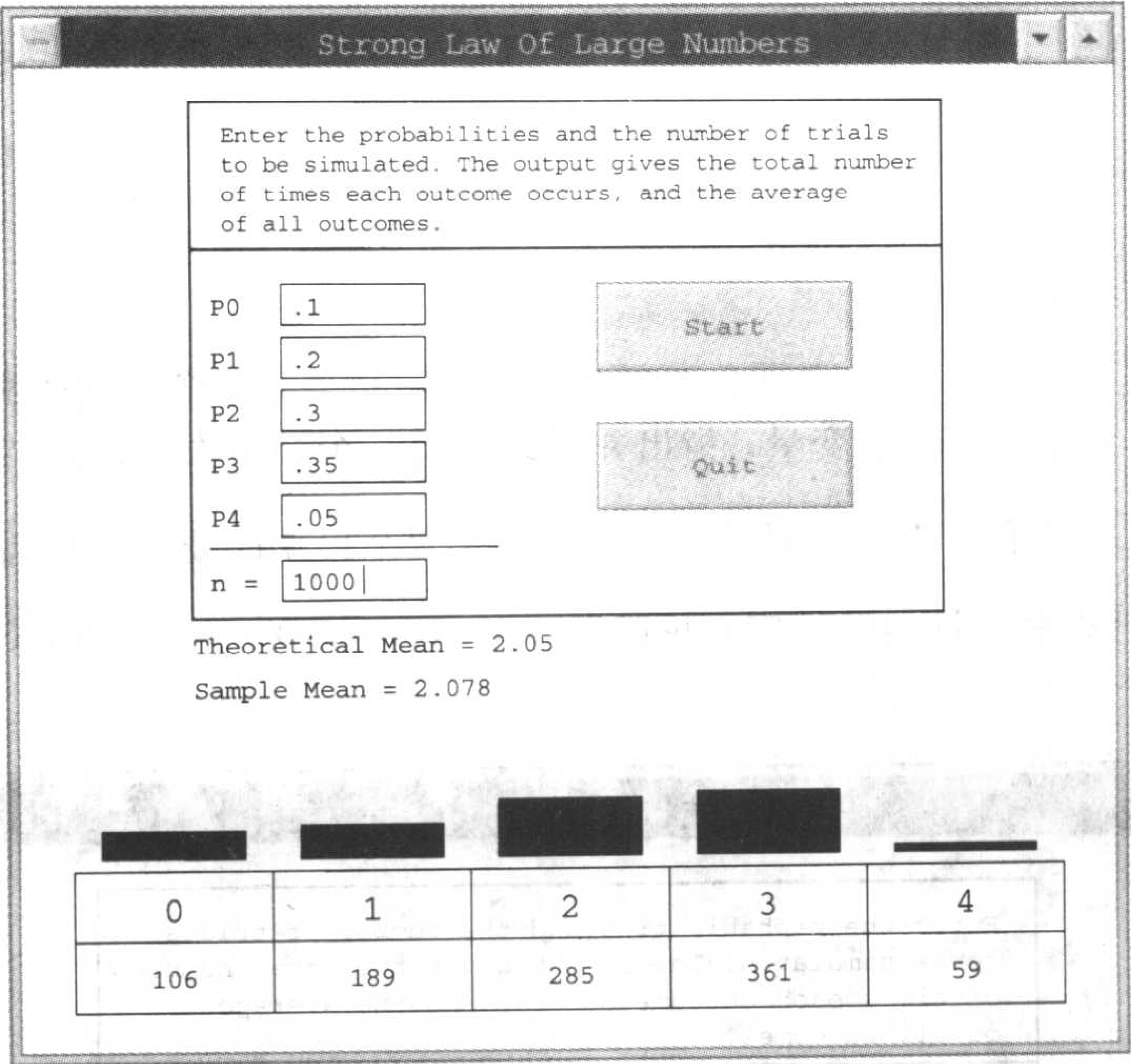
结论得证. ■

对取值为 0，1，2，3，4 的独立同分布随机变量，网站上有两个模块解释了强大数定律. 这两个模块模拟了 n 个这样的随机变量；指出并描绘了每一个结果发生次数的比例以及样本均值 $\sum_{i=1}^{n}X_i/n$. 与上述图表的形式略有不同的是，当使用这两个模块时要键入所需的概率和 n 的期望值. 图 8-2 给出了当概率质量函数给定，并且(a) $n=100$，(b) $n=1\ 000$，(c) $n=10\ 000$ 时的一个模拟结果.

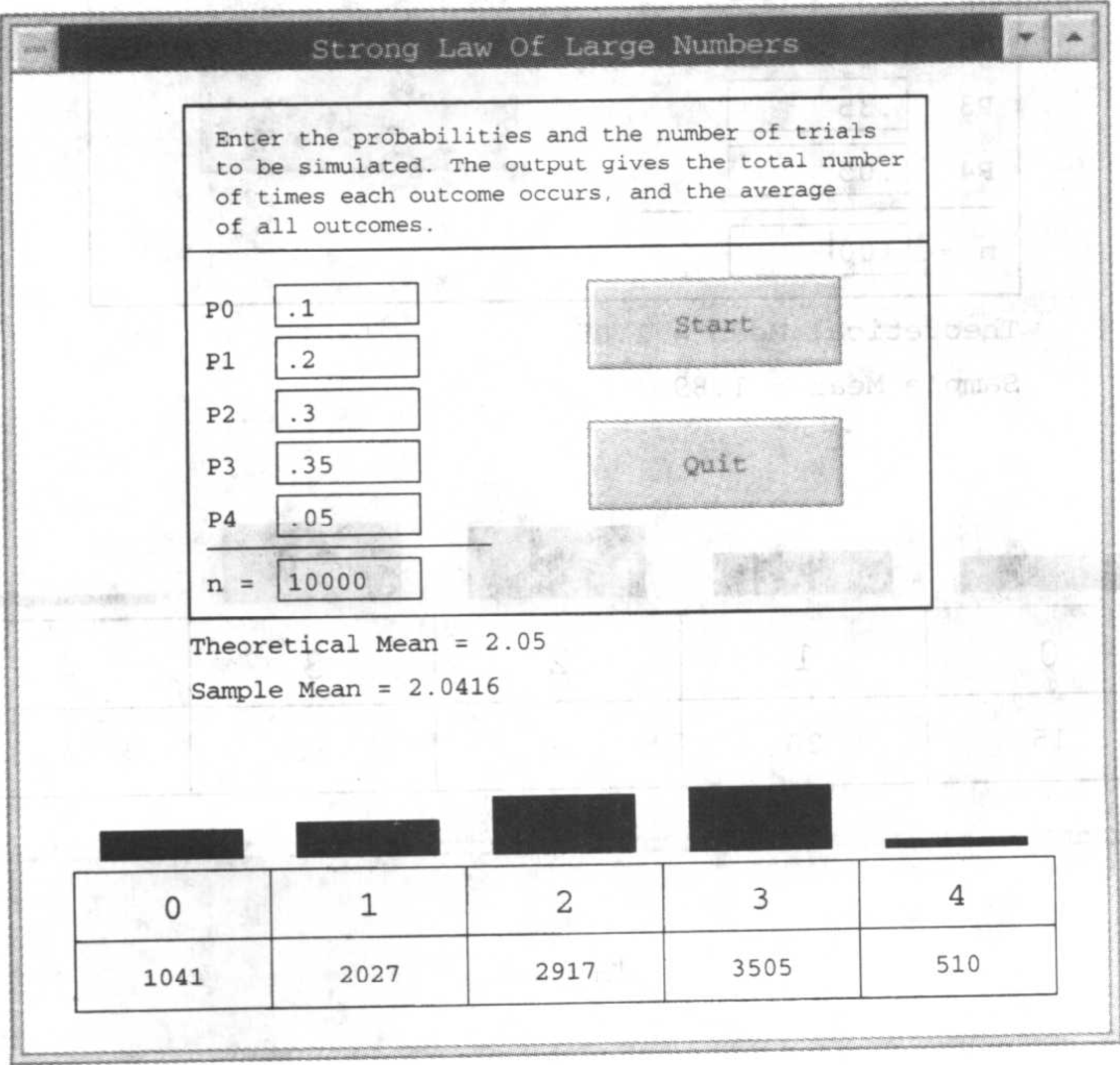


a)

图 8-2


Strong Law Of Large Numbers
Enter the probabilities and the number of trials to be simulated. The output gives the total number of times each outcome occurs, and the average of all outcomes.
P0 .1
P1 .2
P2 .3
P3 .35
P4 .05
n = 1000
Start
Quit
Theoretical Mean = 2.05
Sample Mean = 2.078
0 1 2 3 4
106 189 285 361 59

b)


Strong Law Of Large Numbers
Enter the probabilities and the number of trials to be simulated. The output gives the total number of times each outcome occurs, and the average of all outcomes.
P0 .1
P1 .2
P2 .3
P3 .35
P4 .05
n = 10000
Start
Quit
Theoretical Mean = 2.05
Sample Mean = 2.0416
0 1 2 3 4
1041 2027 2917 3505 510

c)

图 8-2 （续）

有许多初学的学生对弱大数定律与强大数定律之间的区别分辨不清. 弱大数定律指出，对任一固定的、大的数 n^*，$(X_1+\cdots+X_{n^*})/n^*$ 很可能靠近 μ. 然而，这并不是说，对一切大于 n^* 的 n，$(X_1+\cdots+X_n)/n$ 一定接近 μ. 因此 $|(X_1+\cdots+X_n)/n-\mu|$ 的较大的值，仍然有可能出现无穷多次(尽管可能性很小). 而强大数定律指出，这种情况不会发生. 特别是，对任一正数 ε，以概率 1 有

$$\left|\sum_1^n \frac{X_i}{n}-\mu\right|$$

大于 ε 的次数是有限的.

在其中的随机变量为伯努利随机变量的特殊情况下，强大数定律，最初由法国数学家博雷尔(Borel)所证明. 定理 4.1 所示的强大数定律的一般形式，是俄国数学家柯尔莫哥洛夫(A. N. Kolmogorov)证明的.

8.5 其他不等式

有时会遇到这种情况，我们感兴趣的是得到形如 $P\{X-\mu\geqslant a\}$ 的概率的上界，其中 a 为正数，但我们所知道的仅仅是 X 的分布的均值 $\mu=E[X]$ 和方差 $\sigma^2=\mathrm{Var}(X)$. 当然，因 $X-\mu\geqslant a>0$ 蕴涵着 $|X-\mu|\geqslant a$，由切比雪夫不等式可以得到，当 $a>0$ 时有

$$P\{X-\mu\geqslant a\}\leqslant P\{|X-\mu|\geqslant a\}\leqslant\frac{\sigma^2}{a^2}$$

但是，正如下面的命题所示，我们还可以得到更好的上界.

416 ~ 417

命题 5.1(单边切比雪夫不等式) *设 X 为随机变量，其均值为 0，方差 σ^2 有限，则对任意 $a>0$ 有*

$$P\{X\geqslant a\}\leqslant\frac{\sigma^2}{\sigma^2+a^2}$$

证明 令 $b>0$，因为 $X\geqslant a$ 等价于 $X+b\geqslant a+b$，所以

$$P\{X\geqslant a\}=P\{X+b\geqslant a+b\}\leqslant P\{(X+b)^2\geqslant(a+b)^2\}$$

其中这个不等式的得来是因为 $a+b>0$，$X+b\geqslant a+b$ 隐含着 $(X+b)^2\geqslant(a+b)^2$. 对上式应用马尔可夫不等式得

$$P\{X\geqslant a\}\leqslant\frac{E[(X+b)^2]}{(a+b)^2}=\frac{\sigma^2+b^2}{(a+b)^2}$$

令 $b=\dfrac{\sigma^2}{a}$[这显然是使 $(\sigma^2+b^2)/(a+b)^2$ 最小的 b 值]，命题得证. ■

例 5a 设某工厂一周的产量是一个随机变量，均值为 100，方差为 400，试求本周产量超过 120 的概率的上界.

解 用单边切比雪夫不等式得

$$P\{X\geqslant 120\}=P\{X-100\geqslant 20\}\leqslant\frac{400}{400+(20)^2}=\frac{1}{2}$$

因此本周的产量超过 120 的概率至多为 $\dfrac{1}{2}$.

如果用马尔可夫不等式来求上界，那么得到

$$P\{X\geqslant 120\}\leqslant\frac{E(X)}{120}=\frac{5}{6}$$

这个上界比前一个上界要弱的多. ■

现在假设随机变量 X 的均值为 μ，方差为 σ^2. 因为 $X-\mu$ 与 $\mu-X$ 的均值都是 0，方差都是 418 σ^2，则对任意 $a>0$，由单边切比雪夫不等式有

$$P\{X-\mu\geqslant a\}\leqslant\frac{\sigma^2}{\sigma^2+a^2}$$

并且

$$P\{\mu-X\geqslant a\}\leqslant\frac{\sigma^2}{\sigma^2+a^2}$$

于是得到以下推论.

推论 5.1 设 $E[X]=\mu$，$\text{Var}(X)=\sigma^2$，则对任意 $a>0$ 有

$$P\{X\geqslant\mu+a\}\leqslant\frac{\sigma^2}{\sigma^2+a^2}$$

$$P\{X\leqslant\mu-a\}\leqslant\frac{\sigma^2}{\sigma^2+a^2}$$

例 5b 100 位男士与 100 位女士被随机分为 100 对，两人一对. 求至多有 30 组一男一女组成的对的概率的上界.

解 任意对男士从 1 到 100 进行编号，对 $i=1, 2, \cdots, 100$，令

$$X_i=\begin{cases}1 & \text{若第 } i \text{ 个男士与女士配对}\\ 0 & \text{其他}\end{cases}$$

这样，男女配对的对数 X 可以表示为

$$X=\sum_{i=1}^{100}X_i$$

因为第 i 个男士等可能地与其余 199 人配对，其中有 100 位女士，因此

$$E[X_i]=P\{X_i=1\}=\frac{100}{199}$$

类似地，对 $i\neq j$，有

419 $$E[X_iX_j]=P\{X_i=1, X_j=1\}=P\{X_i=1\}P\{X_j=1\mid X_i=1\}=\frac{100}{199}\,\frac{99}{197}$$

其中 $P\{X_j=1\mid X_i=1\}=99/197$，因为给定了第 i 个男士与女士配对，那么第 j 个男士就等可能地与其余的 197 人配对，其中有 99 位女士. 因此可得

$$E[X]=\sum_{i=1}^{100}E[X_i]=(100)\frac{100}{199}\approx 50.25$$

$$\begin{aligned}\text{Var}(X)&=\sum_{i=1}^{100}\text{Var}(X_i)+2\sum_{i<j}\sum\text{Cov}(X_i, X_j)\\&=100\,\frac{100}{199}\,\frac{99}{199}+2\binom{100}{2}\left[\frac{100}{199}\,\frac{99}{197}-\left(\frac{100}{199}\right)^2\right]\approx 25.126\end{aligned}$$

由切比雪夫不等式得

$$P\{X\leqslant 30\}\leqslant P\{|X-50.25|\geqslant 20.25\}\leqslant\frac{25.126}{(20.25)^2}\approx 0.061$$

因此只有不到 6%的机会使最多有 30 位男士与女士配对. 然而，可以用单边切比雪夫不等式使

上述界更精确，得

$$P\{X \leqslant 30\} = P\{X \leqslant 50.25 - 20.25\} \leqslant \frac{25.126}{25.126 + (20.25)^2} \approx 0.058$$ ■

当随机变量 X 的矩母函数已知时，还可求得 $P\{X \geqslant a\}$ 更精确的界. 令

$$M(t) = E[e^{tX}]$$

为随机变量 X 的矩母函数，则对 $t>0$，

$$P\{X \geqslant a\} = P\{e^{tX} \geqslant e^{ta}\} \leqslant E[e^{tX}]e^{-ta} \qquad \text{由马尔可夫不等式}$$

类似地，对 $t<0$，

$$P\{X \leqslant a\} = P\{e^{tX} \geqslant e^{ta}\} \leqslant E[e^{tX}]e^{-ta}$$

由此可得下述被称为切尔诺夫(Chernoff)界的不等式. [420]

命题 5.2(切尔诺夫界)　对任意 $t>0$ 有 $P\{X \geqslant a\} \leqslant e^{-ta}M(t)$.

对任意 $t<0$ 有 $P\{X \leqslant a\} \leqslant e^{-ta}M(t)$.

因为对任意正的或负的 t 值，切尔诺夫界都成立，用使 $e^{-ta}M(t)$ 最小的 t 值可得 $P\{X \geqslant a\}$ 最好的界.

例 5c(标准正态随机变量的切尔诺夫界)　设 Z 为标准正态随机变量，则其矩母函数为 $M(t)=e^{t^2/2}$，从而对任意 $t>0$，$P\{Z \geqslant a\}$ 的切尔诺夫界为

$$P\{Z \geqslant a\} \leqslant e^{-ta} e^{t^2/2}$$

因为使 $e^{-ta} e^{t^2/2}$ 最小与使 $t^2/2 - ta$ 最小的 $t(t>0)$ 相同，均为 $t=a$，从而对 $a>0$ 可得

$$P\{Z \geqslant a\} \leqslant e^{-a^2/2}$$

类似地，对 $a<0$ 有

$$P\{Z \leqslant a\} \leqslant e^{-a^2/2}$$ ■

例 5d(泊松随机变量的切尔诺夫界)　若 X 是一个参数为 λ 的泊松随机变量，那么它的矩母函数是 $M(t)=e^{\lambda(e^t-1)}$. 因此 $P\{X \geqslant i\}$ 的切尔诺夫界是

$$P\{X \geqslant i\} \leqslant e^{\lambda(e^t-1)} e^{-it} \qquad t>0$$

最小化上式右边等价于最小化 $\lambda(e^t-1)-it$，通过积分得到当 $e^t=i/\lambda$ 时达到最小值. 若 $i/\lambda>1$，t 的最小值是正的. 因此，假设 $i>\lambda$ 并且在切尔诺夫界中令 $e^t=i/\lambda$ 得到

$$P\{X \geqslant i\} \leqslant e^{\lambda(i/\lambda-1)} \left(\frac{\lambda}{i}\right)^i$$

或等价地，

$$P\{X \geqslant i\} \leqslant \frac{e^{-\lambda}(e\lambda)^i}{i^i}$$ ■

例 5e　有一个赌徒在每局赌博中都等可能地、独立地赢 1 美元或输 1 美元. 即，若 X_i 是该赌徒在第 i 局中赢得的钱数，则 X_i 相互独立，且

$$P\{X_i = 1\} = P(X_i = -1) = \frac{1}{2}$$ [421]

令 $S_n = \sum_{i=1}^{n} X_i$ 表示 n 局后该赌徒的赢钱数. 我们将用到 $P(S_n \geqslant a)$ 的切尔诺夫界. 首先，注意到 X_i 的矩母函数是

$$E[e^{tX}] = \frac{e^t + e^{-t}}{2}$$

现在，用 e^t 和 e^{-t} 的麦克劳林展开式可以得到

$$\begin{aligned} e^t + e^{-t} &= 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots + \left(1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \cdots\right) \\ &= 2\left\{1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \cdots\right\} \\ &= 2\sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} \leqslant 2\sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} \qquad \text{由于}(2n)! \geqslant n!2^n \\ &= 2e^{t^2/2} \end{aligned}$$

因此，

$$E[e^{tX}] \leqslant e^{t^2/2}$$

由于独立随机变量和的矩母函数是各矩母函数的乘积，我们有

$$E[e^{tS_n}] = (E[e^{tX}])^n \leqslant e^{nt^2/2}$$

由上述结果及切尔诺夫界得到

$$P\{S_n \geqslant a\} \leqslant e^{-ta} e^{nt^2/2} \qquad t > 0$$

使得上式右端最小的 t 值即为最小化 $nt^2/2 - ta$ 的 t 值，此值为 $t = a/n$. 假设 $a > 0$(从而最小化的 t 值是正的)并在上述不等式中令 $t = a/n$ 得到

$$P\{S_n \geqslant a\} \leqslant e^{-a^2/2n} \qquad a > 0$$

例如，由此不等式可得

$$P\{S_{10} \geqslant 6\} \leqslant e^{-36/20} \approx 0.165\ 3$$

而精确值则为

$$\begin{aligned} P\{S_{10} \geqslant 6\} &= P\{\text{赌徒在前 10 局中至少赢 8 局}\} \\ &= \frac{\binom{10}{8} + \binom{10}{9} + \binom{10}{10}}{2^{10}} = \frac{56}{1\ 024} \approx 0.054\ 7 \end{aligned}$$

422

■

下面是一个与其说与概率有关，不如说与数学期望有关的不等式. 在叙述之前，我们需要引进下面的定义.

定义 二次可微的实值函数 $f(x)$ 称为凸的，如果对一切 x，$f''(x) \geqslant 0$；类似地，$f(x)$ 称为凹的，如果对一切 x，$f''(x) \leqslant 0$.

凸函数的一些例子是，$f(x) = x^2$，$f(x) = e^{ax}$，$f(x) = -x^{\frac{1}{n}}$，$x \geqslant 0$. 如果 $f(x)$ 是凸的，则 $g(x) = -f(x)$ 是凹的，反之亦然.

命题 5.3(詹森(Jensen)不等式) 设 $f(x)$ 为凸函数，则有

$$E[f(X)] \geqslant f(E[X])$$

只要上式中的数学期望存在且有限.

证明 将 $f(x)$ 在 $\mu = E[X]$ 附近展开成泰勒级数，得

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + \frac{f''(\xi)(x - \mu)^2}{2}$$

其中 ξ 是 x 与 μ 之间的某个数. 由于 $f''(\xi)\geqslant 0$，我们得到

$$f(x)\geqslant f(\mu)+f'(\mu)(x-\mu)$$

因此

$$f(X)\geqslant f(\mu)+f'(\mu)(X-\mu)$$

两边取数学期望得

$$E[f(X)]\geqslant f(\mu)+f'(\mu)E[X-\mu]=f(\mu)$$

不等式得证. ■

例 5f 一位投资者面临下述选择：她可以将所有钱投入随机回报为 X（X 的均值为 m）的风险资产中；也可以将钱投入以概率 1 有回报 m 的无风险资产中. 假设她的决定建立在最大化 $u(R)$ 的期望值上，其中 R 是她的回报，u 是她的效用函数. 由詹森不等式，若 u 是凹函数，那么 $E[u(X)]\leqslant u(m)$，因此无风险投资更可取；若 u 是凸函数，则 $E[\mu(X)]\geqslant\mu(m)$，那么风险投资更可取. ■ 423

8.6 用泊松随机变量逼近独立伯努利随机变量之和的误差概率界

本节我们将建立由同均值的泊松随机变量逼近独立伯努利随机变量之和的误差概率界. 假设我们想要估计各具有均值 p_1，p_2，…，p_n 的独立伯努利随机变量的和. 首先，令 Y_1，…，Y_n 是一个独立的泊松随机变量序列，Y_i 的均值为 p_i，我们将构造一个参数分别为 p_1，p_2，…，p_n 的独立伯努利随机变量序列 X_1，X_2，…，X_n，使得

$$P\{X_i\neq Y_i\}\leqslant p_i^2\qquad \text{对任意 } i$$

令 $X=\sum_{i=1}^{n}X_i$，$Y=\sum_{i=1}^{n}Y_i$，我们将利用上述事实得到

$$P\{X\neq Y\}\leqslant\sum_{i=1}^{n}p_i^2$$

最后，我们将指出上述不等式隐含着对任意一个实数集 A，有

$$|P\{X\in A\}-P\{Y\in A\}|\leqslant\sum_{i=1}^{n}p_i^2$$

由于 X 是独立的伯努利随机变量之和，Y 是一个泊松随机变量，从而由上面的不等式可以得到需要的界.

为了说明上式是怎样得到的，令 $Y_i(i=1,\cdots,n)$ 是均值为 p_i 的独立泊松随机变量. 现在令 U_1，U_2，…，U_n 是与每个 Y_i 独立的随机变量，并且满足

$$U_i=\begin{cases}0 & \text{以概率}(1-p_i)e^{p_i}\\ 1 & \text{以概率 } 1-(1-p_i)e^{p_i}\end{cases}$$

上面的简洁定义是为了在 $(1-p_i)e^{p_i}\leqslant 1$ 的假设下利用不等式

$$e^{-p}\geqslant 1-p$$

现在定义随机变量 $X_i(i=1,\cdots,n)$ 为

$$X_i=\begin{cases}0 & \text{如果 } Y_i=U_i=0\\ 1 & \text{其他}\end{cases}$$

注意到

$$P\{X_i=0\}=P\{Y_i=0\}P\{U_i=0\}=e^{-p_i}(1-p_i)e^{p_i}=1-p_i$$
$$P\{X_i=1\}=1-P\{X_i=0\}=p_i$$

424
若 X_i 等于 0，那么 Y_i 必定等于 0（由 X_i 的定义）. 因此，我们可以看到

$$\begin{aligned} P\{X_i \neq Y_i\} &= P\{X_i = 1, Y_i \neq 1\} = P\{Y_i = 0, X_i = 1\} + P\{Y_i > 1\} \\ &= P\{Y_i = 0, U_i = 1\} + P\{Y_i > 1\} \\ &= e^{-p_i}[1 - (1 - p_i)e^{p_i}] + 1 - e^{-p_i} - p_i e^{-p_i} \\ &= p_i - p_i e^{-p_i} \\ &\leqslant p_i^2 \qquad (\text{由于 } 1 - e^{-p} \leqslant p) \end{aligned}$$

令 $X = \sum_{i=1}^{n} X_i$，$Y = \sum_{i=1}^{n} Y_i$. 注意到 X 是独立的伯努利随机变量之和，Y 是泊松随机变量，其均值 $E[Y] = E[X] = \sum_{i=1}^{n} p_i$. 还注意到 $X \neq Y$ 蕴涵着某个 i 有 $X_i \neq Y_i$，因此

$$\begin{aligned} P\{X \neq Y\} &\leqslant P\{\text{对某个 } i,\ X_i \neq Y_i\} \\ &\leqslant \sum_{i=1}^{n} P\{X_i \neq Y_i\} \qquad (\text{布尔不等式}) \\ &\leqslant \sum_{i=1}^{n} p_i^2 \end{aligned}$$

对任意事件 B，令 I_B 表示事件 B 的示性变量，定义为

$$I_B = \begin{cases} 1 & \text{若 } B \text{ 发生} \\ 0 & \text{否则} \end{cases}$$

注意到对任意实数集 A，有

$$I_{\{X \in A\}} - I_{\{Y \in A\}} \leqslant I_{\{X \neq Y\}}$$

上式之所以成立是由于示性变量只能取 0 或 1，左边只有当 $I_{\{X \in A\}} = 1$ 且 $I_{\{Y \in A\}} = 0$ 时取 1. 这表明 $X \in A$ 且 $Y \notin A$，这意味着 $X \neq Y$，因此右边也等于 1. 两边取数学期望得到

$$P\{X \in A\} - P\{Y \in A\} \leqslant P\{X \neq Y\}$$

交换 Y 和 X，同理可以得到

$$P\{Y \in A\} - P\{X \in A\} \leqslant P\{X \neq Y\}$$

因此我们可以得到

$$|P\{X \in A\} - P\{Y \in A\}| \leqslant P\{X \neq Y\}$$

于是，我们就证明了对 $\lambda = \sum_{i=1}^{n} p_i$，有

425

$$\left| P\left\{ \sum_{i=1}^{n} X_i \in A \right\} - \sum_{i \in A} \frac{e^{-\lambda} \lambda^i}{i!} \right| \leqslant \sum_{i=1}^{n} p_i^2$$

注 当所有的 p_i 都等于 p 时，X 是一个二项随机变量，那么由上面可知，对任意非负整数集 A，有

$$\left| \sum_{i \in A} \binom{n}{i} p^i (1-p)^{n-i} - \sum_{i \in A} \frac{e^{-np}(np)^i}{i!} \right| \leqslant np^2$$

小结

本章我们介绍了两个有用的概率界——马尔可夫不等式和切比雪夫不等式. 马尔可夫不等式是关于非负随机变量的，它指出，对此种类型的 X 有

$$P\{X \geqslant a\} \leqslant \frac{E[X]}{a}$$

其中，a 是任意正数. 切比雪夫不等式是马尔可夫不等式的一个简单推论，它指出，若 X 的均值为 μ，方差为 σ^2，那么对任意的正数 k，有

$$P\{|X-\mu|\geqslant k\sigma\}\leqslant\frac{1}{k^2}$$

概率论中两个最为重要的理论结果是中心极限定理和强大数定律. 两者都是关于独立同分布的随机变量序列的. 中心极限定理表明，若随机变量具有有限的均值 μ 和有限的方差 σ^2，那么当 n 很大时，它们中前 n 个的和近似服从均值为 $n\mu$、方差为 $n\sigma^2$ 的正态分布. 即，若 $X_i(i\geqslant 1)$ 是满足条件的序列，那么中心极限定理表明，对任意的实数 a，有

$$\lim_{n\to\infty}P\left\{\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\leqslant a\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a}\mathrm{e}^{-x^2/2}\,\mathrm{d}x$$

强大数定律只要求序列中的随机变量具有有限的均值 μ. 该定律表明，当 n 趋于 ∞ 时，它们中前 n 个的平均值以概率 1 收敛到 μ. 这表明，若用 A 表示某个独立重复 n 次的试验的特定事件，那么试验的结果为 A 的极限频率以概率 1 等于 $P(A)$. 因此，若我们"以概率 1"接受这种解释，理解为"必然"，就得到了长期的相对频率的概率解释的证明. [426]

习题

1. 设 X 为随机变量，其均值和方差都等于 20. 对 $P\{0<X<40\}$ 有什么结论?
2. 某教授根据以往的经验知道，一个学生在期末考试中的成绩是均值等于 75 的随机变量.
   (a) 假设这位教授知道该学生成绩的方差为 25，试给出此学生的成绩将超过 85 分的概率的上界.
   (b) 你对这个学生将取得 65 分到 85 分之间的概率有什么结论?
   (c) 不用中心极限定理，求应有多少如上的学生参加考试，才能保证他们的平均分数在 70 分到 80 分之间的概率至少是 0.9?
3. 利用中心极限定理解习题 2 的(c).
4. 设 $X_1, X_2, \cdots, X_{20}$ 是均值为 1 的独立泊松随机变量.
   (a) 用马尔可夫不等式求 $P\left\{\sum_1^{20}X_i>15\right\}$ 的上界.
   (b) 用中心极限定理求 $P\left\{\sum_1^{20}X_i>15\right\}$ 的近似值.
5. 50 个数字被舍入为距它们最近的整数，并对它们求和. 若单个舍入误差服从$(-0.5, 0.5)$上的均匀分布，则所求和与精确和相差超过 3 的概率是多少?
6. 连续地掷一个骰子直到掷的点数之和超过 300，问至少需要掷 80 次的概率是多少?
7. 某人有 100 个灯泡，它们的寿命是均值为 5 小时的指数随机变量且相互独立. 若一个灯泡坏掉后立即换上另一个，525 小时后仍有灯泡在工作的概率是多少?
8. 在习题 7 中假设换一个灯泡所用的时间服从$(0, 0.5)$上的均匀分布. 550 小时后所有灯泡都坏掉的概率是多少?
9. 若 X 是参数为$(n, 1)$的 Γ 随机变量，为使

$$P\left\{\left|\frac{X}{n}-1\right|>0.01\right\}<0.01$$

   n 应该是多大?
10. 国内工程师认为桥的某一段能够免于结构性损害所能承受的重量 W(单位：1000 磅[⊖])服从均值为 400、标准差为 40 的正态分布. 假设轿车的重量(单位：1000 磅)是一个均值为 3、标准差为 0.3 的随机

[⊖] 磅的单位符号是 lb，1 lb=0.453 592 37 kg. ——编辑注

变量. 为使桥受到结构性损害的概率超过 0.1，应有多少辆轿车在桥上?

11. 许多人认为股票市场中某公司股票价格的日改变量是一个均值为 0、方差为 σ^2 的随机变量. 即，若 Y_n 表示第 n 天股票的价格，那么

$$Y_n = Y_{n-1} + X_n \qquad n \geqslant 1$$

其中 X_1，X_2，…是独立同分布的随机变量，均值为 0，方差为 σ^2. 假设今日股票价格为 100. 若 $\sigma^2=1$，对于股票价格在 10 天后将超过 105 的概率，你有什么结论?

427

12. 有 100 个组件会连续地被投入使用. 即，组件 1 首先投入使用，坏掉后换上组件 2，再坏掉后换上组件 3，以此类推. 若第 i 个组件的寿命服从均值为 $10+i/10(i=1, \cdots, 100)$ 的指数分布，估计所有组件的总寿命将超过 1200 小时的概率. 当第 i 个组件的寿命服从 $(0, 20+i/5)(i=1, \cdots, 100)$ 上的均匀分布时，重复上述问题.

13. 由某教员提供的学生成绩的均值为 74，标准差为 14. 该教员欲分别对两个班进行两场考试，其中一个班 25 人，另一个班 64 人.
    (a) 估计 25 人的那个班的平均成绩超过 80 分的概率.
    (b) 对 64 人的那个班重复(a).
    (c) 估计人数多的班级比人数少的班级的平均成绩多 2.2 分的概率.
    (d) 估计人数少的班级比人数多的班级的平均成绩多 2.2 分的概率.

14. 某组件对电力系统的运行是至关重要的，因此坏掉后必须马上更换. 如果该类型组件的平均寿命是 100 小时，标准差是 30 小时，必须储备多少这种组件才能使系统至少以 0.95 的概率在接下来的 2 000 小时内连续运行?

15. 某保险公司有 10 000 名汽车保险人. 每位汽车保险人的年期望索赔为 240 美元，标准差为 800 美元. 估计一年的总索赔额超过 270 万美元的概率.

16. 在男女配对数(近似)为正态分布的假设下，重做例 5b. 此假设合理吗?

17. 在已知学生测验成绩的方差为 25 的情况下，重复习题 2 的(a)部分.

18. 某湖中有 4 种不同类型的鱼. 假设每条捕到的鱼都等可能地属于这些类型. 令 Y 表示每种类型的鱼至少捕得一条所需的捕鱼数量.
    (a) 给定一个区间 (a, b)，使得 $P\{a \leqslant Y \leqslant b\} \geqslant 0.90$.
    (b) 利用单边切比雪夫不等式，我们至少应计划捕多少条鱼才能以 90%的把握断定每种类型的鱼至少捕得一条?

19. 若 X 是均值为 25 的非负随机变量，对于下列各项有何结论?
    (a) $E[X^3]$；　(b) $E[\sqrt{X}]$；
    (c) $E[\log X]$；　(d) $E[e^{-X}]$.

20. 令 X 是一个非负随机变量，证明

$$E[X] \leqslant (E[X^2])^{1/2} \leqslant (E[X^3])^{1/3} \leqslant \cdots$$

21. 若允许投资者将他的钱分割，并将占总体比例为 $\alpha(0<\alpha<1)$ 的那部分钱投入风险资产，其余的钱投入无风险资产，那么例 5f 中的结果会不会改变? 分割后的回报将是 $R=\alpha X+(1-\alpha)m$.

22. 令 X 是均值为 20 的泊松随机变量.
    (a) 利用马尔可夫不等式得到

$$p = P\{X \geqslant 26\}$$

    的一个上界.
    (b) 利用单边切比雪夫不等式得到 p 的一个上界.
    (c) 利用切尔诺夫界得到 p 的一个上界.
    (d) 利用中心极限定理估计 p.

(e) 通过合适的程序确定 p.

428

理论练习

1. 若 X 的方差为σ^2，那么方差的正平方根 σ 称为标准差. 若 X 的均值为μ，标准差为 σ，证明

$$P\{|X-\mu|\geqslant k\sigma\}\leqslant\frac{1}{k^2}$$

2. 若 X 的均值为μ，标准差为σ，称比率 $r\equiv|\mu|/\sigma$ 为 X 的信噪比测度，其解释为：X 可以表示为 $X=\mu+(X-\mu)$，其中 μ 为信号，$X-\mu$ 为噪声. 若定义$|(X-\mu)/\mu|\equiv D$ 为 X 与其信号(均值)μ 的相对偏差，那么对 $\alpha>0$，证明

$$P\{D\leqslant\alpha\}\geqslant 1-\frac{1}{r^2\alpha^2}$$

3. 计算下述随机变量的信噪比测度，即$|\mu|/\sigma$，其中 $\mu=E[X]$，$\sigma^2=\mathrm{Var}(X)$.
   (a) 均值为 λ 的泊松随机变量； (b) 参数为 n 和 p 的二项随机变量；
   (c) 均值为 $1/p$ 的几何随机变量；(d) (a, b)上的均匀随机变量；
   (e) 均值为 $1/\lambda$ 的指数随机变量；(f) 参数为 μ，σ^2 的正态随机变量.
4. 令 $Z_n(n\geqslant1)$是一个随机变量序列，c 是一个常数使得对任意 $\varepsilon>0$，当 $n\to\infty$时，有 $P\{|Z_n-c|>\varepsilon\}\to0$，证明对任意有界连续函数 g，有

$$E[g(Z_n)]\to g(c)，当 n\to\infty$$

5. 令 $f(x)$是定义在 $0\leqslant x\leqslant1$ 上的连续函数. 考虑函数

$$B_n(x)=\sum_{k=0}^{n}f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}$$

(称为伯恩斯坦(Bernstein)多项式)并证明

$$\lim_{n\to\infty}B_n(x)=f(x)$$

提示：令 X_1，X_2，…是均值为 x 的独立伯努利随机变量. 利用事实(用理论练习 4 中的结果)

$$B_n(x)=E\left[f\left(\frac{X_1+\cdots+X_n}{n}\right)\right]$$

由于可以证明 $B_n(x)$收敛到 $f(x)$对 x 是一致的，从而以上提供了分析中著名的魏尔斯特拉斯(Weierstrass)定理的概率证明，也就是定义在闭区间上的任意连续函数可由一个多项式任意逼近.

6. (a)令 X 是一个离散型随机变量，取值为 1，2，…. 若 $P\{X=k\}$对 $k=1, 2, \cdots$是非增的，证明

$$P\{X=k\}\leqslant 2\,\frac{E[X]}{k^2}$$

(b) 令 X 是一个非负的连续型随机变量，有非增的密度函数. 证明

$$f(x)\leqslant\frac{2E[X]}{x^2}\qquad 对所有\ x>0$$

429

7. 设掷一个均匀骰子 100 次，令 X_i 表示第 i 次掷出的点数，求

$$P\left\{\prod_{i=1}^{100}X_i\leqslant a^{100}\right\},\ 1<a<6$$

的近似值.

8. 解释一下为什么一个参数为(t, λ)的 Γ 随机变量当 t 很大时有一个正态近似.
9. 设抛一枚均匀硬币 1 000 次，如果前 100 次都出现正面，在余下的 900 次中，你期望正面出现的比例占多大？试评述下述说法："强大数定律小得无法补救了".
10. 若 X 是一个均值为 λ 的泊松随机变量，证明对 $i<\lambda$，

$$P\{X \leqslant i\} \leqslant \frac{e^{-\lambda}(e\lambda)^i}{i^i}$$

11. 令 X 是一个参数为 n 和 p 的二项随机变量. 证明对 $i>np$:

(a) 当 t 满足 $e^t=\frac{iq}{(n-i)p}$, $q=1-p$ 时, $\min_{t>0} e^{-ti}E[e^{tX}]$达到.

(b) $P\{X\geqslant i\}\leqslant\frac{n^n}{i^i(n-i)^{n-i}}p^i(1-p)^{n-i}$.

12. 由标准正态随机变量 Z 的切尔诺夫界得到 $P\{Z>a\}\leqslant e^{-a^2/2}$, $a>0$. 通过考虑 Z 的密度函数, 证明不等式的右端可以减少$\frac{1}{2}$. 即证明

$$P\{Z > a\} \leqslant \frac{1}{2}e^{-a^2/2} \qquad a > 0$$

13. 若 $E[X]<0$ 并且 $\theta\neq 0$ 使得 $E[e^{\theta X}]=1$, 证明 $\theta>0$.

自测题与练习

1. 某汽车经销商的周销售量是一个随机变量, 期望值是 16. 给出下述概率的上界:

(a) 下周的销售量超过 18; (b) 下周的销售量超过 25.

2. 假设上题中周销售量的方差是 9.

(a) 对下周的销售量在 10 到 22 之间的概率给出一个下界.

(b) 对下周的销售量超过 18 的概率给出一个上界.

3. 若

$$E[X] = 75 \quad E[Y] = 75 \quad \mathrm{Var}(X) = 10 \quad \mathrm{Var}(Y) = 12 \quad \mathrm{Cov}(X, Y) = -3$$

对下述概率给出一个上界:

(a) $P\{|X-Y|>15\}$; (b) $P\{X>Y+15\}$; (c) $P\{Y>X+15\}$.

4. 假设工厂 A 的日产量是一个随机变量, 均值为 20, 标准差为 3. 工厂 B 的日产量也是一个随机变量, 均值为 18, 标准差为 6. 假设它们相互独立, 对今日的产量 B 厂超过 A 厂的概率给出一个上界.

430

5. 某型号的元件在坏掉之前的工作时间是一个随机变量, 概率密度函数是

$$f(x) = 2x \qquad 0 < x < 1$$

一旦元件坏掉后立刻换上另一个同类型的新元件. 若令 X_i 表示投入使用的第 i 个元件的寿命, 那么 $S_n = \sum_{i=1}^{n} X_i$ 表示第 n 个元件坏掉时的时间. 毁坏发生时的长期比 r 的定义如下:

$$r = \lim_{n\to\infty} \frac{n}{S_n}$$

假设 $X_i(i\geqslant 1)$是相互独立的, 确定 r.

6. 在自测题 5 中, 手头需要有多少元件才能以 90%的把握断定库存将会维持 35 天?

7. 机器维修需要两个独立的步骤, 第一个步骤所需时间是一个均值为 0.2 小时的指数随机变量, 第二个步骤所需时间是另一个独立的均值为 0.3 小时的指数随机变量. 若维修人员有 20 台机器需要修理, 估计所有工作将在 8 小时内完成的概率.

8. 在每局赌博中, 赌徒以 0.7 的概率输掉 1 美元, 以 0.2 的概率输掉 2 美元, 以 0.1 的概率赢得 10 美元. 估计赌徒在下第 100 次注后会输的概率.

9. 在自测题 7 中确定 t, 使得维修人员在时间 t 内完成 20 项工作的概率大约是 0.95.

10. 某烟草公司声明在他们生产的卷烟中, 每支尼古丁的含量是一个均值为 2.2 mg, 标准差为 0.3 mg 的随机变量. 但是, 随机抽取的 100 支卷烟中尼古丁的平均含量为 3.1 mg. 若该公司的声明是可信的,

估计平均值至少为 3.1 的概率.

11. 某诊所在给定的一天等可能地有 2，3 或 4 位志愿医生来提供服务. 无论在这一天有多少志愿医生，来找这些医生看病的病人数是均值为 30 的独立随机变量. 令 X 表示这一天在该诊所看病的病人.
    (a) 求 $E[X]$
    (b) 求 $\mathrm{Var}(X)$.
    (c) 用一个标准正态概率分布表估计 $P\{X>65\}$.

431

第 9 章　概率论的其他主题

9.1　泊松过程

在定义泊松过程之前，我们先来回忆一下：如果$\lim_{h\to 0} f(h)/h=0$，我们称函数 f 是 $o(h)$. 也就是说，对任意小的 h 值，$f(h)$相对 h 而言更小. 假设事件发生的时间点是随机的，并设 $N(t)$ 表示在时间段$[0, t]$内发生的事件数，则随机变量的集合$\{N(t), t\geqslant 0\}$称为具有强度 $\lambda(\lambda>0)$的泊松过程，如果满足下列条件：

(i) $N(0)=0$.

(ii) 不相交时间段内发生的事件数独立.

(iii) 在给定时间段内发生的事件数的分布只依赖于时间段的长度，与其位置无关.

(iv) $P\{N(h)=1\}=\lambda h+o(h)$.

(v) $P\{N(h)\geqslant 2\}=o(h)$.

条件(i)表示过程从 0 时刻开始. 条件(ii)即**独立增量**(independent increment)假设，表示在时刻 t 时发生的事件数[即 $N(t)$]与时刻 t 到时刻 $t+s$ 之间发生的事件数[即 $N(t+s)-N(t)$]独立. 条件(iii)即**平稳增量**(stationary increment)假设，表示对所有的 t 值，$N(t+s)-N(t)$的概率分布是相同的.

在第 4 章中我们曾阐述过一个结论：根据泊松分布是二项分布的极限，我们得出上述条件隐含着 $N(t)$服从均值为 λt 的泊松分布. 现在我们用另外一种方法证明这个结果.

引理 1.1　*对强度为 λ 的泊松过程，有*

$$P\{N(t)=0\}=e^{-\lambda t}$$

432

证明　令 $P_0(t)=P\{N(t)=0\}$，用下述方式导出关于 $P_0(t)$的一个微分方程：

$$\begin{aligned} P_0(t+h) &= P\{N(t+h)=0\} = P\{N(t)=0, N(t+h)-N(t)=0\} \\ &= P\{N(t)=0\}P\{N(t+h)-N(t)=0\} = P_0(t)[1-\lambda h+o(h)] \end{aligned}$$

其中后两个等式是由于条件(ii)以及条件(iv)和(v)蕴涵着 $P\{N(h)=0\}=1-\lambda h+o(h)$，因此，

$$\frac{P_0(t+h)-P_0(t)}{h}=-\lambda P_0(t)+\frac{o(h)}{h}$$

令 $h\to 0$ 得

$$P_0'(t)=-\lambda P_0(t)$$

或等价地，有

$$\frac{P_0'(t)}{P_0(t)}=-\lambda$$

积分得

$$\log P_0(t)=-\lambda t+c$$

或

$$P_0(t)=Ke^{-\lambda t}$$

因为 $P_0(0)=P\{N(0)=0\}=1$，所以

$$P_0(t)=e^{-\lambda t}$$



对一个泊松过程，令 T_1 表示第一个事件发生的时间. 进而，对 $n>1$，令 T_n 表示从第 $n-1$ 个事件发生到第 n 个事件发生的时间间隔，序列 $\{T_n, n=1, 2, \cdots\}$ 称为时间间隔序列(sequence of interarrival time). 例如，若 $T_1=5$，$T_2=10$，则泊松过程的第一个事件在时刻5发生，第二个事件在时刻15发生.

下面我们来看 T_n 的分布，首先注意到事件 $\{T_1>t\}$ 发生，当且仅当在 $[0, t]$ 内泊松过程没有事件发生，因此

$$P\{T_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}$$

从而，T_1 服从均值为 $1/\lambda$ 的指数分布，并且

433

$$P\{T_2 > t\} = E[P\{T_2 > t \mid T_1\}]$$

然而，

$$\begin{aligned} P\{T_2 > t \mid T_1 = s\} &= P\{\text{在}(s, s+t]\text{内发生 0 个事件} \mid T_1 = s\} \\ &= P\{\text{在}(s, s+t]\text{内发生 0 个事件}\} = e^{-\lambda t} \end{aligned}$$

其中后两个等式是由于独立性和平稳增量的假设. 因此，由以上分析我们得出 T_2 也是均值为 $1/\lambda$ 的指数随机变量，并且 T_2 和 T_1 独立，重复同样的讨论得到命题1.1.

命题 1.1 *$T_1, T_2, \cdots$ 是独立的均值为 $1/\lambda$ 的指数随机变量.*

另一个令人感兴趣的量是 S_n，它表示第 n 个事件的到达时刻，也称为直到第 n 个事件出现的等待时间. 容易看出

$$S_n = \sum_{i=1}^{n} T_i \qquad n \geqslant 1$$

因此由命题1.1和5.6.1节的结果可知 S_n 服从参数为 n 和 λ 的 Γ 分布. 也就是说，S_n 的概率密度函数为

$$f_{S_n}(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} \qquad x \geqslant 0$$

下面证明 $N(t)$ 是均值为 λt 的泊松随机变量.

定理 1.1 *对强度为 λ 的泊松随机过程，有*

$$P\{N(t) = n\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

证明 因为泊松过程的第 n 个事件在 t 时刻或 t 时刻以前发生，当且仅当到时刻 t 为止发生的事件数至少为 n，即

434

$$N(t) \geqslant n \Leftrightarrow S_n \leqslant t$$

因此

$$\begin{aligned} P\{N(t) = n\} &= P\{N(t) \geqslant n\} - P\{N(t) \geqslant n+1\} = P\{S_n \leqslant t\} - P\{S_{n+1} \leqslant t\} \\ &= \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} dx - \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^n}{n!} dx \end{aligned}$$

但是根据分部积分公式 $\int u dv = uv - \int v du$，令 $u = e^{-\lambda x}$，$dv = \lambda[(\lambda x)^{n-1}/(n-1)!]dx$，得

$$\int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} dx = e^{-\lambda t} \frac{(\lambda t)^n}{n!} + \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^n}{n!} dx$$

定理得证. ■

9.2 马尔可夫链

考虑随机变量序列 X_0，X_1，…，并假定这些随机变量的可能值的集合为$\{0, 1, \cdots, M\}$. 把 X_n 理解为某系统在时刻 n 的状态是有益的，按照这种理解，若 $X_n=i$，则称这个系统在时刻 n 处于状态 i. 称一个随机变量序列构成一个马尔可夫链(Markov chain)，如果每当此系统处于状态 i，存在某个固定的概率(称为 P_{ij})使得系统在下一个时刻以概率 P_{ij} 处于状态 j. 也就是说，对所有的 i_0，…，i_{n-1}，i，j，都有

$$P\{X_{n+1}=j\,|\,X_n=i,\ X_{n-1}=i_{n-1},\ \cdots,\ X_1=i_1,\ X_0=i_0\}=P_{ij}$$

值 $P_{ij}(0\leqslant i\leqslant M,\ 0\leqslant j\leqslant N)$称为马尔可夫链的**转移概率**(transition probability)，转移概率满足(为什么?)

$$P_{ij}\geqslant 0\quad \sum_{j=0}^{M}P_{ij}=1\quad i=0,1,\cdots,M$$

为方便起见，将转移概率 P_{ij} 排成一个方阵：

$$\begin{vmatrix} P_{00} & P_{01} & \cdots & P_{0M} \\ P_{10} & P_{11} & \cdots & P_{1M} \\ \vdots & \vdots & \vdots & \vdots \\ P_{M0} & P_{M1} & \cdots & P_{MM} \end{vmatrix}$$

称为**转移概率矩阵**.

从理论上讲，已知转移概率矩阵和 X_0 的分布，就能计算出与此链有关的所有概率. 例如，X_0，…，X_n 的联合概率质量函数为

$$\begin{aligned}
&P\{X_n=i_n,\ X_{n-1}=i_{n-1},\ \cdots,\ X_1=i_1,\ X_0=i_0\}\\
&\quad=P\{X_n=i_n\,|\,X_{n-1}=i_{n-1},\ \cdots,\ X_0=i_0\}P\{X_{n-1}=i_{n-1},\ \cdots,\ X_0=i_0\}\\
&\quad=P_{i_{n-1},i_n}P\{X_{n-1}=i_{n-1},\ \cdots,\ X_0=i_0\}
\end{aligned}$$

435

连续地重复这一推理可得上述概率等于

$$P_{i_{n-1},i_n}P_{i_{n-2},i_{n-1}}\cdots P_{i_1,i_2}P_{i_0,i_1}P\{X_0=i_0\}$$

例 2a 假定明天是否下雨只取决于今天是否下雨. 进一步假设，若今天下雨，则明天也下雨的概率是 α，若今天不下雨，则明天下雨的概率是 β.

若约定下雨时系统处于状态 0，不下雨时处于状态 1，则上述系统是一个二状态马尔可夫链，其转移概率矩阵为

$$\begin{vmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{vmatrix}$$

也就是说，$P_{00}=\alpha=1-P_{01}$，$P_{10}=\beta=1-P_{11}$. ■

例 2b 考虑一赌徒，他在每一局中以概率 p 赢 1 美元而以概率 $1-p$ 输 1 美元，如果假定赌徒手中钱数达到 0 美元或 M 美元时，他就停止赌博，那么赌徒手中的钱数序列是一个马尔可夫链，其转移概率为

$$P_{i,i+1}=p=1-P_{i,i-1}\qquad i=1,\cdots,M-1\qquad P_{00}=P_{MM}=1$$ ■

例 2c 物理学家 P. 埃伦费斯特与 T. 埃伦费斯特(Ehrenfest)研究过分子运动的理想模型，有 M 个分子被放到两个箱中，在每一个时刻，随机地选择一个分子，将它从所在的箱中取出放到另外一个箱内. 若以 X_n 表示第 n 次移动分子以后第一个箱内的分子数，那么$\{X_0$，…，

$X_n\}$ 是一个马尔可夫链，其转移概率为

$$P_{i,i+1}=\frac{M-i}{M} \qquad 0\leqslant i\leqslant M$$

$$P_{i,i-1}=\frac{i}{M} \qquad 0\leqslant i\leqslant M$$

$$P_{ij}=0 \qquad 若 |j-i|>1$$ ■

因此，对一个马尔可夫链而言，P_{ij} 表示系统处于状态 i 下一步转移到状态 j 的概率. 我们还可以定义二步转移概率 $P_{ij}^{(2)}$，即系统由状态 i 出发经两步转移到状态 j 的概率，也就是说，

$$P_{ij}^{(2)}=P\{X_{m+2}=j\,|\,X_m=i\}$$

436 从 P_{ij} 计算 $P_{ij}^{(2)}$ 如下：

$$P_{ij}^{(2)}=P\{X_2=j\,|\,X_0=i\}=\sum_{k=0}^{M}P\{X_2=j,X_1=k\,|\,X_0=i\}$$

$$=\sum_{k=0}^{M}P\{X_2=j\,|\,X_1=k,X_0=i\}P\{X_1=k\,|\,X_0=i\}=\sum_{k=0}^{M}P_{kj}P_{ik}$$

一般地，用 $P_{ij}^{(n)}$ 表示 n 步转移概率，即

$$P_{ij}^{(n)}=P\{X_{n+m}=j\,|\,X_m=i\}$$

命题 2.1 说明如何计算 $P_{ij}^{(n)}$，称之为查普曼-科尔莫戈罗夫方程(Chapman-Kolmogorov equation).

命题 2.1(查普曼-科尔莫戈罗夫方程)

$$P_{ij}^{(n)}=\sum_{k=0}^{M}P_{ik}^{(r)}P_{kj}^{(n-r)} \qquad 对一切\ 0<r<n$$

证明

$$P_{ij}^{(n)}=P\{X_n=j\,|\,X_0=i\}=\sum_{k}P\{X_n=j,X_r=k\,|\,X_0=i\}$$

$$=\sum_{k}P\{X_n=j\,|\,X_r=k,X_0=i\}P\{X_r=k\,|\,X_0=i\}=\sum_{k}P_{kj}^{(n-r)}P_{ik}^{(r)}$$ ■

例 2d(随机运动) 一个有可数无限状态空间的马尔可夫链的例子是随机运动(random walk)，它是一个粒子沿着一维坐标轴移动的轨迹. 假设此粒子在每一个点上以概率 p 向右移动一步，以概率 $1-p$ 向左移动一步，即假设此粒子的轨迹是一个马尔可夫链，其转移概率为

$$P_{i,i+1}=p=1-P_{i,i-1} \qquad i=0,\pm 1,\cdots$$

若粒子的初始位置是 i，那么 n 步转移以后它到达位置 j 的概率就是粒子向右移 $(n-i+j)/2$ 步，向左移动 $n-[(n-i+j)/2]=(n+i-j)/2$ 步的概率. 因为向右移每一步的概率都是 p，并

437 独立于其他任何一步，因此上述概率就是二项分布

$$P_{ij}^{n}=\binom{n}{(n-i+j)/2}p^{(n-i+j)/2}(1-p)^{(n+i-j)/2}$$

其中，当 x 是大于等于 n 的非负整数时，$\binom{n}{x}$ 等于 0. 上式可写成

$$P_{i,i+2k}^{2n}=\binom{2n}{n+k}p^{n+k}(1-p)^{n-k} \qquad k=0,\pm 1,\cdots,\pm n$$

$$P_{i,i+2k+1}^{2n+1}=\binom{2n+1}{n+k+1}p^{n+k+1}(1-p)^{n-k} \qquad k=0,\pm 1,\cdots,\pm n,-(n+1)$$ ■

尽管 $P_{ij}^{(n)}$ 表示条件概率，但是可以通过加上初始位置这个条件，将表达式转换为非条件概率. 例如，

$$P\{X_n = j\} = \sum_i P\{X_n = j \mid X_0 = i\} P\{X_0 = i\} = \sum_i P_{ij}^{(n)} P\{X_0 = i\}$$

对很大的马尔可夫链来说，当 $n \to \infty$ 时，$P_{ij}^{(n)}$ 趋于一个只依赖于 j 的数 Π_j. 也就是说，当 n 很大时，经过 n 步转移到 j 的概率近似等于 Π_j，而与初始状态无关. 马尔可夫链具有这个性质的充分条件是：对某个 $n>0$，有

$$P_{ij}^{(n)} > 0 \qquad 对一切\ i,\ j = 0,\ 1,\ \cdots,\ M \tag{2-1}$$

满足式(2-1)的马尔可夫链称为是*遍历的*(ergodic). 由命题 2.1 得

$$P_{ij}^{(n+1)} = \sum_{k=0}^{M} P_{ik}^{(n)} P_{kj}$$

对遍历链而言，令 $n \to \infty$，得

$$\Pi_j = \sum_{k=0}^{M} \Pi_k P_{kj} \tag{2-2}$$

而且，由于 $1 = \sum_{j=0}^{M} P_{ij}^{(n)}$，再令 $n \to \infty$，我们还可以得到

$$\sum_{j=0}^{M} \Pi_j = 1 \tag{2-3}$$

438

事实上，可以证明 $\Pi_j (0 \leqslant j \leqslant M)$ 是式(2-2) 与式(2-3) 的唯一非负解. 这些结果都包括在定理 2.1 中，我们不加证明地叙述如下：

定理 2.1 *对遍历马尔可夫链而言，*

$$\Pi_j = \lim_{n \to \infty} P_{ij}^{(n)}$$

存在，并且 $\Pi_j (0 \leqslant j \leqslant M)$ 是方程

$$\Pi_j = \sum_{k=0}^{M} \Pi_k P_{kj} \qquad \sum_{j=0}^{M} \Pi_j = 1$$

的唯一非负解.

例 2e 考虑例 2a，其中我们假设，如果今天下雨，那么明天下雨的概率是 α；如果今天不下雨，那么明天下雨的概率是 β. 由定理 2.1 得，下雨和不下雨的极限概率 Π_0 和 Π_1 满足

$$\Pi_0 = \alpha \Pi_0 + \beta \Pi_1 \qquad \Pi_1 = (1-\alpha)\Pi_0 + (1-\beta)\Pi_1 \qquad \Pi_0 + \Pi_1 = 1$$

解之得

$$\Pi_0 = \frac{\beta}{1+\beta-\alpha} \qquad \Pi_1 = \frac{1-\alpha}{1+\beta-\alpha}$$

例如，若 $\alpha = 0.6$，$\beta = 0.3$，那么第 n 天下雨的极限概率是 $\Pi_0 = 3/7$. ■

Π_j 也等于马尔可夫链处于状态 $j (j = 0, \cdots, M)$ 的时间长度比例. 凭直觉应该是这样的：令 P_j 表示马尔可夫链处于位置 j 的时间长度比例(用强大数定律可以证明，对于遍历链而言，此长度比例存在并且是常数)，因为马尔可夫链在状态 k 的概率是 P_k，并且以概率 P_{kj} 转移到状态 j，因此马尔可夫链从状态 k 进入状态 j 的概率等于 $P_k P_{kj}$. 对所有的 k 相加，马尔可夫链进入状态 j 的时间比例满足

$$P_j = \sum_k P_k P_{kj}$$

439

显然

$$\sum_j P_j = 1$$

因此，由定理2.1有$\Pi_j(j=0,\cdots,M)$是上述方程的唯一解，并且长度比例的解释对不遍历链来说一般也是成立的.

例2f 考虑在例2c中，第一个箱中有j个分子，其中$j=0,\cdots,M$，由定理2.1，这些量是下述这些方程的唯一解：

$$\Pi_0 = \Pi_1 \times \frac{1}{M}$$

$$\Pi_j = \Pi_{j-1} \times \frac{M-j+1}{M} + \Pi_{j-1} \times \frac{j+1}{M} \quad j=1,\cdots,M$$

$$\Pi_M = \Pi_{M-1} \times \frac{1}{M}$$

$$\sum_{j=0}^{M} \Pi_j = 1$$

然而，很容易验证

$$\Pi_j = \binom{M}{j}\left(\frac{1}{2}\right)^M \qquad j=0,\cdots,M$$

也满足上述方程，从而这些就是马尔可夫链处于每一个状态的时间长度比例(见习题11，解释如何猜出上述解). ■

9.3 意外、不确定性与熵

考虑进行一个试验可能发生的事件E，当得知E事实上发生了的时候，我们会有多大的意外呢？看起来，假定这个消息所引起的意外程度的大小取决于E的概率是合乎情理的. 例如，如果试验是掷一对骰子，那么当E表示点数之和为偶数这一事件时(这时E有概率1/2)，我们听到E已发生不会感到非常意外；但是若E表示点数之和是12这一事件时(此时其概率是1/36)，我们得知E已发生一定会感到非常意外.

本节我们将试图给出意外概念的量化. 首先，约定某人在获悉事件E已发生所感到的"意外"只依赖于E的概率，并以$S(p)$表示概率为p的事件发生时所引起的意外. 将分两步来确定$S(p)$的函数形式：先确定$S(p)$应满足的一组合理的条件，再证明这些公理要求$S(p)$必须具备某个特定形式. 假设$S(p)$对一切$0<p\leqslant 1$是有定义的，而对于$p=0$的事件则无定义.

440

第一个条件正好叙述了这样一个明显的事实：当听到某必定发生的事件已发生时，我们不会感到意外.

公理1

$$S(1) = 0$$

第二个条件是说，比较起来越不可能发生的事件若发生了，它引起的意外也更大.

公理2 *$S(p)$是一个严格递减的函数，即若$p<q$，则$S(p)>S(q)$.*

第三个条件从数学上阐述了这样一个事实：我们直观上希望p的一个微小改变对应$S(p)$的一个微小改变.

公理3 *$S(p)$是p的连续函数.*

为引出最后一个条件，考虑两个相互独立的事件E与F，它们的概率分别是$P(E)=p$，

$P(F)=q$，因为 $P(EF)=pq$，那么事件 E 和 F 都发生所引起的意外就是 $S(pq)$. 假设先已知 E 发生，然后才知道 F 也发生了. 因为 $S(p)$是 E 发生引起的意外，故当被告知 F 也已发生时，增添的意外应是 $S(pq)-S(p)$. 但因 F 与 E 相互独立，E 的发生并不改变 F 发生的概率，因此增添的意外正好等于 $S(q)$. 这一推理启发我们给出最后一个条件.

公理 4

$$S(pq) = S(p) + S(q) \qquad 0 < p \leqslant 1,\ 0 < q \leqslant 1$$

至此已为定理 3.1 做好了准备，这个定理给出了 $S(p)$的构造.

441

定理 3.1 *若 $S(\cdot)$满足公理 1 至 4，那么*

$$S(p) = -C\log_2 p$$

其中 C 是任意正整数.

证明 由公理 4 得

$$S(p^2) = S(p) + S(p) = 2S(p)$$

由归纳法可得

$$S(p^m) = mS(p) \tag{3-1}$$

其次，因为对任意正整数 n，$S(p)=S(p^{1/n}\cdots p^{1/n})=nS(p^{1/n})$，所以有

$$S(p^{1/n}) = \frac{1}{n}S(p) \tag{3-2}$$

于是，由式(3-1)和式(3-2)得

$$S(p^{m/n}) = mS(p^{1/n}) = \frac{m}{n}S(p)$$

这等价于对一切正有理数 x，有

$$S(p^x) = xS(p) \tag{3-3}$$

但是由 S 的连续性(公理 3)知，式(3-3)对所有的非负数 x 均成立(试推导这一结果).

现对任意的 p，$0<p\leqslant 1$，令 $x=-\log_2 p$，则 $p=\left(\frac{1}{2}\right)^x$. 再由式(3-3)得

$$S(p) = S\left(\left(\frac{1}{2}\right)^x\right) = xS\left(\frac{1}{2}\right) = -C\log_2 p$$

其中 $C=S\left(\frac{1}{2}\right)$，由公理 1 和 2 得 $C=S\left(\frac{1}{2}\right)>S(1)=0$. ■

通常取 $C=1$，这时称意外以比特(bit，二进制数字 binary digit 的缩写)为单位表出.

现考虑一随机变量 X，它必须取 x_1，…，x_n 中的某个值，且取各值的概率依次为 p_1，…，p_n. 因为 X 取值 x_i 引起的意外等于 $-\log p_i$[⊖]，因此 X 的值引起的意外程度的期望值为

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i$$

442

在信息论中，量 $H(X)$称为随机变量 X 的熵(entropy).(当 $p_i=0$ 时，令 $0\log 0=0$.)可以证明(留为习题)，当所有的 p_i 相等时，$H(X)$取得最大值.(这是否符合直观?)

因为$H(X)$表示某人获悉X的值时所感到意外的平均值，故它还可以理解为关于X取值

⊖ 在本章的余下部分，我们把 $\log_2 x$ 简记为 $\log x$，而用 $\ln x$ 表示 $\log_e x$.

的不确定性(uncertainty). 事实上，在信息论中，$H(x)$被解释为观察 X 的值时所获得的平均信息(information)量. 因此，X 引起的平均意外、X 的不确定性以及由 X 产生的平均信息量不过是从三个稍微不同的观点得到的同一个概念.

现考虑两个随机变量 X 与 Y，它们分别取值 $x_1, \cdots, x_n$ 与 $y_1, \cdots, y_m$，其联合质量函数是

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}$$

从而用 $H(X, Y)$表示随机向量(X, Y)值的不确定性是

$$H(X, Y) = -\sum_i \sum_j p(x_i, y_j)\log p(x_i, y_j)$$

假设 Y 的观测值是 y_j，在此条件下遗留在 X 中的不确定性就是

$$H_{Y=y_j}(X) = -\sum_i p(x_i | y_j)\log p(x_i | y_j)$$

其中

$$p(x_i | y_j) = P\{X = x_i | Y = y_j\}$$

因此观测到 Y 以后，遗留在 X 中的平均不确定性就是

$$H_Y(X) = \sum_j H_{Y=y_j}(X) p_Y(y_j)$$

其中

$$p_Y(y_j) = P\{Y = y_j\}$$

命题 3.1 将 $H(X, Y)$表为 $H(Y)$与 $H_Y(X)$，它表明 X 与 Y 取值的不确定性，等于 Y 的不确定性加上 Y 被观测到以后遗留在 X 中的平均不确定性.

命题 3.1

443

$$H(X, Y) = H(Y) + H_Y(X)$$

证明 由等式 $p(x_i, y_j) = p_Y(y_j) p(x_i | y_j)$得

$$\begin{aligned}
H(X, Y) &= -\sum_i \sum_j p(x_i, y_j)\log p(x_i, y_j) \\
&= -\sum_i \sum_j p_Y(y_j) p(x_i | y_j)[\log p_Y(y_j) + \log p(x_i | y_j)] \\
&= -\sum_j p_Y(y_j)\log p_Y(y_j) \sum_i p(x_i | y_j) - \sum_j p_Y(y_j) \sum_i p(x_i | y_j)\log p(x_i | y_j) \\
&= H(Y) + H_Y(X)
\end{aligned}$$

■

信息论的一个基本结论是，当随机变量 Y 被观测到以后，随机变量 X 的不确定性平均来说减小了. 在证明这一结论之前，需引入下述引理，其证明留作练习.

引理 3.1

$$\ln x \leqslant x - 1 \qquad x > 0$$

仅当 $x=1$ 时，等号成立.

定理 3.2

$$H_Y(X) \leqslant H(X)$$

当且仅当 X 与 Y 独立时，等号成立.

证明

$$H_Y(X) - H(X) = -\sum_i \sum_j p(x_i | y_j)\log[p(x_i | y_j)]p(y_j) + \sum_i \sum_j p(x_i, y_j)\log p(x_i)$$

$$= \sum_i \sum_j p(x_i, y_j) \log\left[\frac{p(x_i)}{p(x_i|y_j)}\right]$$

$$\leqslant \log e \sum_i \sum_j p(x_i, y_j)\left[\frac{p(x_i)}{p(x_i|y_j)} - 1\right] \quad \text{由引理 3.1}$$

444

$$= \log e\left[\sum_i \sum_j p(x_i)p(y_j) - \sum_i \sum_j p(x_i, y_j)\right] = \log e[1-1] = 0 \quad \blacksquare$$

9.4 编码论与熵

假设在 A 地观测某离散型随机变量 X 的值，然后通过一通信线路用 0 与 1 两个字符传送到 B 地. 为此，首先应把 X 的每一个可能值编码成由 0 与 1 组成的序列. 为了避免含糊不清，通常要求不能有这样的编码序列，即由一个较短的编码序列添加几个字符而得到.

例如，如果 X 有四个可能值 x_1，x_2，x_3，x_4，那么可以形成一种编码方案

$$\begin{aligned} x_1 &\leftrightarrow 00 \\ x_2 &\leftrightarrow 01 \\ x_3 &\leftrightarrow 10 \\ x_4 &\leftrightarrow 11 \end{aligned} \tag{4-1}$$

也就是说，若 $X=x_1$，那么传送到 B 地的电文是 00，若 $X=x_2$，那么把 01 传送到 B 地，等等. 另一种可采纳的编码方案是：

$$\begin{aligned} x_1 &\leftrightarrow 0 \\ x_2 &\leftrightarrow 10 \\ x_3 &\leftrightarrow 110 \\ x_4 &\leftrightarrow 111 \end{aligned} \tag{4-2}$$

然而，像

$$\begin{aligned} x_1 &\leftrightarrow 0 \\ x_2 &\leftrightarrow 1 \\ x_3 &\leftrightarrow 00 \\ x_4 &\leftrightarrow 01 \end{aligned}$$

这样的编码方案不能用，因为 x_3，x_4 的编码序列都可由延长 x_1 的编码序列而得到.

我们的目标之一是，寻求使由 A 地传送到 B 地需要的比特（二进制数字）数的期望最小的编码. 例如，若

445

$$\begin{aligned} P\{X = x_1\} &= 1/2 \\ P\{X = x_2\} &= 1/4 \\ P\{X = x_3\} &= 1/8 \\ P\{X = x_4\} &= 1/8 \end{aligned}$$

那么式(4-2)给出的编码平均要传送 $1/2(1)+1/4(2)+1/8(3)+1/8(3)=1.75$ 个比特；然而式(4-1)给出的编码则平均要传送 2 个比特. 因此，对以上概率分布而言，按式(4-2)编码比按式(4-1)编码更有效.

这就提出了如下问题：对一个给定的随机向量 X，编码方案能达到的最大效率是多少？答案是：对任何编码而言，需传送的平均比特数不小于 X 的熵. 在信息论中，这一结果称为无噪

声编码定理(noiseless coding theorem). 为了证明它，需要引入引理 4.1.

引理 4.1 设 X 可能取值为 $x_1, \cdots, x_N$，那么，当把 X 的值编码为比特序列(其中任意一个序列都不是其他序列的延长)时，为使各比特序列的长度依次为 $n_1, \cdots, n_N$，其充分必要条件是

$$\sum_{i=1}^{N}\left(\frac{1}{2}\right)^{n_i} \leqslant 1$$

证明 对固定的 N 个正整数 $n_1, \cdots, n_N$，设 w_j 表示等于 j 的 n_i 的个数，其中 $j=1, \cdots$. 为了能使值 x_i 的编码长度为 n_i 个比特，其中 $i=1, \cdots, N$，显然必须有 $w_1 \leqslant 2$. 而且，由于要求每个比特序列不能是另一比特序列的延长，故必须有 $w_2 \leqslant 2^2-2w_1$，(这是因为长度为 2 的比特序列的个数是 2^2，而其中由 w_1 个长度为 1 的比特序列延长而得到的个数为 $2w_1$.)同理可以证明，对 $n=1, \cdots$，必有

$$w_n \leqslant 2^n - w_1 2^{n-1} - w_2 2^{n-2} - \cdots - w_{n-1} 2 \tag{4-3}$$

事实上，读者稍加思考就会确信，对于使 x_i 的编码长度为 n_i 个比特，其中 $i=1, \cdots, N$，上述条件不仅是必要的而且是充分的.

将不等式(4-3)改写为

$$w_n + w_{n-1}2 + w_{n-2}2^2 + \cdots + w_1 2^{n-1} \leqslant 2^n \qquad n = 1, \cdots$$

两边同除以 2^n 就得到充分必要条件：对一切 n 有

$$\sum_{j=1}^{n} w_j \left(\frac{1}{2}\right)^j \leqslant 1 \tag{4-4}$$

446

然而，因为 $\sum_{j=1}^{n} w_j \left(\frac{1}{2}\right)^j$ 随着 n 的增大而增大，于是当且仅当

$$\sum_{j=1}^{\infty} w_j \left(\frac{1}{2}\right)^j \leqslant 1$$

时式(4-4)成立. 再由定义知 w_j 表示 n_j 中等于 j 的个数，从而

$$\sum_{j=1}^{\infty} w_j \left(\frac{1}{2}\right)^j = \sum_{i=1}^{N}\left(\frac{1}{2}\right)^{n_i}$$ ■

下面证明定理 4.1.

定理 4.1(**无噪声编码定理**) 设 X 分别以概率 $p(x_1), \cdots, p(x_N)$ 取值 $x_1, \cdots, x_N$，则对于任何使 x_i 有 n_i 个比特的 X 的编码，

$$\sum_{i=1}^{N} n_i p(x_i) \geqslant H(X) = -\sum_{i=1}^{N} p(x_i)\log p(x_i)$$

证明 令 $P_i = p(x_i)$，$q_i = 2^{-n_i} / \sum_{j=1}^{N} 2^{-n_j}$，$i = 1, \cdots, N$，则

$$\begin{aligned}
-\sum_{i=1}^{N} P_i \log\left(\frac{P_i}{q_i}\right) &= -\log e \sum_{i=1}^{N} P_i \ln\left(\frac{P_i}{q_i}\right) \\
&= \log e \sum_{i=1}^{N} P_i \ln\left(\frac{q_i}{P_i}\right) \\
&\leqslant \log e \sum_{i=1}^{N} P_i \left(\frac{q_i}{P_i} - 1\right) \qquad \text{由引理 3.1} \\
&= 0 \qquad \text{因为} \sum_{i=1}^{N} P_i = \sum_{i=1}^{N} q_i = 1
\end{aligned}$$

因此

$$-\sum_{i=1}^{N}P_i\log P_i\leqslant-\sum_{i=1}^{N}P_i\log q_i=\sum_{i=1}^{N}n_iP_i+\log\left(\sum_{j=1}^{N}2^{-n_j}\right)$$

$$\leqslant\sum_{i=1}^{N}n_iP_i\qquad\text{由引理 4.1}$$

■ 447

例 4a 考虑随机变量 X，其概率质量函数为

$$p(x_1)=\frac{1}{2}\qquad p(x_2)=\frac{1}{4}\qquad p(x_3)=p(x_4)=\frac{1}{8}$$

因为

$$H(X)=-[1/2\log 1/2+1/4\log 1/4+1/4\log 1/8]$$
$$=1/2+2/4+3/4=1.75$$

所以，由定理 4.1 知没有比

$$x_1\leftrightarrow 0$$
$$x_2\leftrightarrow 10$$
$$x_3\leftrightarrow 110$$
$$x_4\leftrightarrow 111$$

更好的编码方案. ■

对多数随机向量而言，不存在使传送的比特数的平均值达到其下界 $H(X)$ 的编码，但是，总可以设计一种编码方案，使得它的比特数的平均值在 $H(X)$ 与 $H(X)+1$ 之间. 为证明这一点，以 n_i 表示满足

$$-\log p(x_i)\leqslant n_i<-\log p(x_i)+1$$

的整数. 因为

$$\sum_{i=1}^{N}2^{-n_i}\leqslant\sum_{i=1}^{N}2^{\log p(x_i)}=\sum_{i=1}^{N}p(x_i)=1$$

从而由引理 4.1，对 $x_i(i=1,\cdots,N)$ 可找到长度为 n_i 的比特序列，此序列的平均长度 $L=\sum_{i=1}^{N}n_ip(x_i)$ 满足

$$-\sum_{i=1}^{N}p(x_i)\log p(x_i)\leqslant L<-\sum_{i=1}^{N}p(x_i)\log p(x_i)+1$$

或者

$$H(X)\leqslant L<H(X)+1$$

例 4b 假定独立地抛一枚硬币 10 次，此硬币出现正面的概率为 p. 试验是在 A 地进行的，而且要把结果传送到 B 地. 试验的结果是一随机向量 $X=(X_1,\cdots,X_{10})$，其中 X_i 取 1 还是 0 依赖于第 i 次抛硬币出现正面或反面而定. 由这一节的结论可知，采用任何一种编码方案所传送的比特数的平均值 L 满足 448

$$H(X)\leqslant L$$

并且至少有一种编码方案满足

$$L\leqslant H(X)+1$$

由于 X_i 相互独立，从而由命题 3.1 和定理 3.2 知

$$H(X)=H(X_1,\cdots,X_n)=\sum_{i=1}^{N}H(X_i)=-10[p\log p+(1-p)\log(1-p)]$$

若 $p=1/2$，则 $H(X)=10$，因此以 X 的真实值作为它的编码是最好的. 也就是说，例如，若前 5 次抛出正面，后 5 次抛出反面，则传送给 B 地的信息应是 1111100000.

然而，若 $p\neq 1/2$，则采用其他编码方案往往更好一些. 例如，若 $p=1/4$，则

$$H(X)=-10(1/4\log 1/4+3/4\log 3/4)=8.11$$

因此，存在一种编码方案，其编码信息的平均长度不大于 9.11.

此时，比这个编码更有效的简单编码是：把 $(X_1,\cdots,X_{10})$ 分成 5 对，每对含两个随机变量，然后把每一对随机变量编码如下：对 $i=1,3,5,7,9$，取

$$X_i=0,X_{i+1}=0\leftrightarrow 0$$
$$X_i=0,X_{i+1}=1\leftrightarrow 10$$
$$X_i=1,X_{i+1}=0\leftrightarrow 110$$
$$X_i=1,X_{i+1}=1\leftrightarrow 111$$

用这种编码方案传送的整个信息比按实际值传送的信息要好.

例如，若得到的结果是 $TTTHHTTTTH$，那么传送的信息是 010110010. 使用此编码传送信息，所需的比特数的平均值是

$$5\left[1\left(\frac{3}{4}\right)^2+2\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)+3\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)+3\left(\frac{1}{4}\right)^2\right]=\frac{135}{16}\approx 8.44$$ ■

到目前为止我们总是假定，从 A 地发出的信息被 B 地无差错地接收. 然而，由于通信系统中有随机干扰，所以免不了产生某些误差. 例如，在 A 地发出的信息是 00101101，但是在 B 地收到的却是 01101101.

449

假设从 A 地发送的每个比特被 B 地正确地接收的概率为 p，且每个比特是否出错独立于其他比特. 这样的通信系统称为二进对称信道(binary symmetric channel). 进一步假设 $p=0.8$ 并且要从 A 地传送到 B 地的信息由大量的比特组成. 因此，由于每个比特出错的概率为 0.20，故若采取直接传送信息的方案，此信息传错的概率是十分高的. 为降低这个概率，可将每个比特重复发送 3 次，而译码则采取多数原则. 也就是说，采用如下方案：

编码	译码
0→000	000, 001, 010, 100 → 0
1→111	111, 110, 101, 011 → 1

注意，若在传送中不多于 1 个差错，则这个比特将被正确地译码. 于是，每个比特出错的概率就降到：$(0.2)^3+3(0.2)^2(0.8)=0.104$，这是一个很大的改进. 事实上，采用多次重复发送与多数原则译码这一方案，可以把比特出错的概率降到我们所需的任意小. 例如，下面的方案会把比特出错的概率降到 0.01 以下：

编　码	译　码
0→17 个 0 的串 1→17 个 1 的串	多数原则

上述这类编码方案的问题是，尽管降低了比特出错的概率，但它是以降低所发送的每个信号的比特的效率为代价的(见表 9-1).

表 9-1　重复传送信息的编码方案

每个比特出错的概率	信号的平均传送效率
0.20	1
0.10	$0.33\left(=\frac{1}{3}\right)$
0.01	$0.06\left(=\frac{1}{17}\right)$

450

实际上，这时读者可能自然会想到，比特出错的概率下降到 0 必然导致所传送的每个信号的比特的效率也下降到 0. 但情况不是这样的，在信息论中有一个著名的结果，即香农(Claude Shannon)的*噪声编码定理*(noisy coding theorem). 在定理 4.2 中叙述这一结果.

定理 4.2(噪声编码定理)　*存在数 C，使得对任何比 C 小的数 R 及任意 $\varepsilon>0$，总存在一种编码-译码方案，以平均传送效率 R 比特发送每个信号，而每个比特出错的概率小于 ε. 这种 C 的最大值 C^* 称为信道容量*[⊖]. *对二进对称信道而言，*

$$C^*=1+p\log p+(1-p)\log(1-p)$$

小结

强度为 λ 的泊松过程是一随机变量的集合 $\{N(t),t\geqslant 0\}$，它是随机发生事件的过程. 例如，$N(t)$ 表示时刻 0 到时刻 t 发生的事件数. 泊松过程的特性如下：

(i) 不相交时间段内发生的事件数独立.

(ii) 在给定时间段内发生的事件数的分布只依赖于时间段的长度.

(iii) 每一个时刻只有一个事件发生.

(iv) 事件发生的强度是 λ.

可以看出，$N(t)$ 是均值为 λt 的泊松随机变量. 另外，若 $T_i(i\geqslant 1)$ 表示两次连续事件之间的时间间隔，那么它们是参数为 λ 的独立指数随机变量.

随机变量序列 $X_n(n\geqslant 0)$ 被称为转移概率为 $P_{i,j}$ 的马尔可夫链，X_n 的每一个值分别取自 $0,\cdots,M$ 中的一个，若对所有的 $n,i_0,\cdots.i_n,i,j$，都有

$$P\{X_{n+1}=j\mid X_n=i,X_{n-1}=i_{n-1},\cdots,X_1=i_1,X_0=i_0\}=P_{i,j}$$

若将 X_n 看成是某系统在时刻 n 的状态，则马尔可夫链就是具有如下特征的一个连续状态序列：无论何时进入独立于任何其他状态的状态 i，然后以概率 $P_{i,j}$ 进入下一个状态 j，这对所有的状态 i,j 都成立. 对许多马尔可夫链而言，在时刻 n 处于状态 j 的概率收敛到一个极限值，

⊖ 用熵解释 C^* 见习题 18.

451 而与初始状态无关. 若用 $\Pi_j(j=0,\cdots,M)$ 表示这些极限概率，则下面的方程有唯一解：

$$\Pi_j=\sum_{i=0}^{M}\Pi_iP_{i,j}\qquad j=0,\cdots,M$$

$$\sum_{j=1}^{M}\Pi_j=1$$

另外，Π_j 也等于马尔可夫链处于状态 j 的时间长度比例.

令 X 是一个以概率$\{p_1,\cdots,p_n\}$取 n 个可能值的随机变量，量

$$H(X)=-\sum_{i=1}^{n}p_i\log_2(p_i)$$

称为 X 的熵. 它可以解释为关于 X 取值的不确定性的平均量，或者观察 X 的值时所获得的平均信息量. 对 X 的二进制编码而言，熵有很重要的含义.

理论练习与习题

1. 顾客以泊松强度 λ 进入银行. 假设第一个小时内有两个顾客进入银行，求以下概率：
   (a) 两人都在前 20 分钟内进入银行.
   (b) 至少一人在前 20 分钟内进入银行.
2. 高速公路上汽车以强度为 $\lambda=3$(单位:分钟)的泊松过程穿过一个固定点，如果 Al 盲目地穿过高速公路，那么他用 s 秒穿过公路而不受伤的概率是多少？(假设有车经过的时候他在高速公路上，就认为他受伤.) 对 $s=2,5,10,20$ 求出概率.
3. 在习题 2 中，假设 Al 足够敏捷，只有一辆车经过的时候他不会受伤，但当他穿过高速公路时遇到两辆或两辆以上的车就会受伤，求他用 s 秒穿过公路而不受伤的概率是多少？对 $s=5,10,20,30$ 求出概率.
4. 3 个白球与 3 个黑球分装在两个箱子中，每箱装 3 个球. 若第一箱有 i 个白球，$i=1,2,3$，我们称系统处于状态 i. 每次从两箱中各取出一球，并把第一箱中取出的球放入第二箱、第二箱中取出的球放入第一箱. 设 X_n 表示这样取 n 次之后系统所处的状态，试求马尔可夫链$\{X_n,n\geqslant 0\}$的转移概率.
5. 考虑例 2a，若今天有一半的可能下雨，求出从现在开始下 3 天雨的概率，其中 $\alpha=0.7,\beta=0.3$.
6. 计算习题 4 中的模型的极限概率.

452

7. 一个转移概率矩阵，如果满足对一切 $j=0,\cdots,M$ 有 $\sum_{i=0}^{M}P_{ij}=1$，则称为重随机的. 若这个马尔可夫链是遍历的，试证 $\Pi_j=1/(M+1)$，$j=0,\cdots,M$.
8. 在某一天，Buffy 可能高兴(c)、一般(s)或沮丧(g). 若他今天高兴，则明天分别以概率 0.7,0.2,0.1 处于状态 c，s，g；若他今天一般，则明天分别以概率 0.4，0.3，0.3 处于状态 c，s，g；若他今天沮丧，则明天分别以概率 0.2,0.4,0.4 处于状态 c，s，g. 求 Buffy 高兴的次数比例是多少？
9. 假设明天是否下雨只依赖于前两天的天气情况，具体来说，假设昨天和今天下雨了，那么明天下雨的概率是 0.8；若昨天下雨了，而今天没有下雨，则明天下雨的概率是 0.3；若今天下雨了，而昨天没有下雨，则明天下雨的概率是 0.4；若今天和昨天都没有下雨，则明天下雨的概率是 0.2. 求下雨天数的比例.
10. 某人每天早上去跑步，当离开家去跑步时，他走前门和走后门是等可能的；类似地，当跑完步回家时，他走前门和走后门也是等可能的. 此人有 5 双跑鞋，他回来时把鞋子脱在任意的门. 如果他离开家去跑步时正好走的那个门没有鞋，他就光脚跑步.
    (a) 把它做成马尔可夫链，给出状态和转移概率.
    (b) 求他光脚跑步的次数比例.

11. 此问题和例 2f 有关.
    (a) 证明 Π_j 的值满足必要条件.
    (b) 对任意给定的分子，它在第一个箱中的(极限)概率是多少?
    (c) 在一个很大的时刻，分子 $j(j\geqslant1)$在第一个箱中的事件（极限上)相互独立，对此你有什么结论?
    (d) 解释所给出的极限概率.
12. 试求掷一对均匀骰子所得点数之和的熵.
13. 若 X 分别以概率 $P_1,\cdots,P_n$ 取某 n 个可能值，试证当 $P_i=1/n(i=1,\cdots,n)$时，$H(X)$达到极大值. 这时 $H(X)$等于多少?
14. 掷一对均匀的骰子，设

$$X=\begin{cases}1 & 若点数之和为 6\\ 0 & 其他\end{cases}$$

并设 Y 等于第一个骰子的点数. 试求(a) $H(Y)$. (b) $H_Y(X)$, (c) $H(X,Y)$.

15. 某硬币抛出正面的概率是 $p=\frac{2}{3}$，现将它抛 6 次，计算这个试验结果的熵.
16. 一随机变量 X 分别以概率 $p(x_i)(i=1,\cdots,n)$取 n 个可能的值 $x_1,\cdots,x_n$，我们试图通过一系列提问来获得 X 的值，对每次提问只回答“是”或“不是”. 例如，我们可以问：“$X=x_1$ 吗?”或“X 等于 x_1 或 x_2 或 x_3 吗?”等等. 关于为获知 X 的值需要提问的平均次数，你有什么结论?
17. 对任意的离散型随机变量 X 与函数 f，试证

$$H(f(X))\leqslant H(X)$$

18. 由 A 地传送一个比特到 B 地，如果 X 表示 A 地发送这个比特的值，而 Y 表示 B 地收到的值，那么称 $H(X)-H_Y(X)$为从 A 地到 B 地的信息传输率. 最大传输率，作为 $P\{X=1\}=1-P\{X=0\}$的函数，称为信道容量. 试证，对于满足 $P\{Y=1|X=1\}=P\{Y=0|X=0\}=p$ 的二进对称信道而言，$P\{X=1\}=\frac{1}{2}$时的信息传输率就是其信道容量，并且它的值等于 $1+p\log p+(1-p)\log(1-p)$.

453

自测题与练习

1. 事件以每小时强度为 $\lambda=3$(单位:小时)的泊松过程发生.
   (a) 早晨 8 点到 10 点没有事件发生的概率是多少?
   (b) 早晨 8 点到 10 点发生的事件数的期望值是多少?
   (c) 下午 2 点之后第 5 个事件期望在什么时刻发生?
2. 顾客以强度为 λ(单位:小时)的泊松过程到达一个指定的零售商店. 假设在第一个小时内有两个顾客到达，求以下概率：
   (a) 两个顾客都在前 20 分钟内到达.
   (b) 在前 30 分钟内至少有一个顾客到达.
3. 路上每 5 辆卡车中就有 4 辆后面跟有汽车，而每 6 辆汽车中就有 1 辆后面跟有卡车，求路上卡车占车辆总数的比例.
4. 一小镇每天的天气被划分为阴、晴、多云但干燥. 若某一天的天气为阴，那么第二天的天气是晴或者多云的概率是一样的；若某一天的天气不是阴，那么第二天的天气会以三分之一的概率保持前一天的天气状态，并且如果改变，它会等概率地变为另外两种天气状态. 长远来看，晴天的概率是多少? 阴天的概率有多大?
5. 假设随机变量 X 分别以概率0.35，0.2，0.2，0.2，0.05 取五个可能值，假设随机变量 Y 分别以概率 0.05，0.35，0.1，0.15，0.35 取五个可能值.

(a) 证明 $H(X)>H(Y)$.

(b) 使用习题 13 的结论给出上述不等式的直观解释.

参考文献

9.1、9.2 节

KEMENY, J., L. SNELL, and A. KNAPP. *Denumerable Markov Chains.* New York: D. Van Nostrand Company, 1966.

PARZEN, E. *Stochastic Processes.* San Francisco: Holden-Day, Inc., 1962.

ROSS, S. M. *Introduction to Probability Models.* 7th ed. San Diego, Calif.: Academic Press, Inc., 2000.

ROSS, S. M. *Stochastic Processes.* 2nd ed. New York: John Wiley & Sons, Inc., 1996.

9.3、9.4 节

ABRAMSON, N. *Information Theory and Coding.* New York: McGraw-Hill Book Company, 1963.

MCELIECE, R. *Theory of Information and Coding.* Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1977.

PETERSON, W., and E. WELDON. *Error Correcting Codes*, 2nd ed. Cambridge, Mass.: The MIT Press, 1972.

454

第10章 模　　拟

10.1 引言

假定用一副牌做单人纸牌游戏(solitaire，类似接龙的扑克牌游戏，一副牌有 52 张，并且有某个固定的游戏规则)，如何求出将获胜的概率呢? 一个办法是，从手中这副纸牌的所有(52)! 种可能排列是等可能的这一合理假设出发，设法确定可以获胜的排列有多少种. 可是，由于(52)! 这个数相当大，而且每一特定的排列是否能最终获胜还取决于做这个游戏的人，因此没有求出可以获胜的排列个数的系统方法，可见这个办法是行不通的.

事实上，要确定单人纸牌游戏获胜的概率，在数学上似乎是难以处理的. 但是，并非一切希望都没有了，因为概率论不仅仅是一个数学分支，而且还属于应用科学的范畴. 而在所有的应用科学中，做试验都是一种有价值的方法. 在这个例子中，试验可采取大量重复这个游戏的方式进行，或者比这更好地，编一个程序用计算机来做这个试验. 重复 n 次试验之后，如果令

$$X_i=\begin{cases}1 & \text{若第 } i \text{ 次获胜}\\ 0 & \text{若第 } i \text{ 次失败}\end{cases}$$

则 $X_i(i=1,\cdots,n)$是独立的伯努利随机变量，其数学期望为

$$E[X_i]=P\{\text{游戏中获胜}\}$$

因此，由强大数定律得到

$$\sum_{i=1}^{n}\frac{X_i}{n}=\frac{\text{获胜的次数}}{\text{游戏总次数}}$$

将以概率 1 收敛到 $P\{$游戏中获胜$\}$. 也就是说，通过大量重复这一试验，我们可以用获胜次数所占的比例作为获胜概率的一个估计值. 这种通过试验来求概率的方法称为模拟(Simulation).

455

为了用计算机进行模拟研究，我们必须能产生一个服从(0,1)上均匀分布的随机变量值；这样的变量称为随机数. 为了产生这样的随机数，大多数计算机都有一个内置的、称为随机数产生器的子程序，其输出为一个伪随机数序列. 对所有实际的应用，这是一序列与取自(0,1)上均匀分布的随机数没有什么区别的数. 大多数随机数产生器都以初值 X_0(称为种子)开始，然后通过指定正整数 a，c 和 m 递归计算，并令

$$X_{n+1}=(aX_n+c)\bmod m \quad n\geqslant 0$$

上式的意思是：aX_n+c 除以 m 所得的余数作为 X_{n+1}的值. 这样每一个 X_n 都是 $0,1,\cdots,m-1$ 中的一个，并且 X_n/m 被作为服从(0,1)上均匀分布的随机变量的近似值. 可以证明，选择合适的 a，c 和 m，上述方法产生的随机数可以看作是服从(0,1)上均匀分布的独立随机变量.

作为模拟中的起点，我们假设从服从(0,1)上均匀分布进行模拟，并且用取自这个分布的独立随机变量作为模拟中的随机数.

在单人纸牌游戏中，我们需要用计算机结束这个开始于一个给定顺序的游戏. 然而，因为初始的顺序被等可能地认为是(52)! 种排列中的任一种，能产生一种随机排列也是必需的. 仅用随机数，下面的算法说明了如何完成上面的想法. 这个算法开始于随机选择一个元素，然后将这个元素放在第 n 个位置；然后再从剩余的元素中随机选择一个元素，并将这个选出的元素放在第 $n-1$ 个位置；等等. 这种方法通过保持这些元素的顺序表，并在表中随机选择一个位置，在剩余的元素中有效地进行了随机选择.

例 1a（产生一个随机排列） 假设我们对产生整数 $1,2,\cdots,n$ 的一个排列（$n!$ 种排列是等可能出现的）感兴趣. 从任一种排列开始，经过 $n-1$ 步，我们将可以做到这一点，在每一步中，互换排列中两个数的位置. 为了明了这个排列，我们令 $X(i)(i=1,2,\cdots,n)$ 表示位置 i 的当前数. 具体的算法如下：

1. 考虑任一种随机排列，令 $X(i)(i=1,2,\cdots,n)$ 表示位置 i 的元素.（例如，可以取 $X(i)=i,i=1,2,\cdots,n$.）

2. 产生一个随机变量 N_n，它等可能地是 $1,2,\cdots,n$ 中的任一个.

3. 互换 $X(N_n)$ 和 $X(n)$ 的值. 固定 $X(n)$ 的值.（例如，假设 $n=4$，开始时 $X(i)=i,i=1,2,3,4$. 如果 $N_4=3$，那么新的排列是 $X(1)=1,X(2)=2$，$X(3)=4,X(4)=3$，并且元素 3 将 [456] 始终保持在位置 4.）

4. 产生一个随机变量 N_{n-1}，它等可能地是 $1,2,\cdots,n-1$ 中的任一个.

5. 互换 $X(N_{n-1})$ 和 $X(n-1)$ 的值.（如果 $N_3=1$，那么新排列是 $X(1)=4$，$X(2)=2,X(3)=1,X(4)=3$.）

6. 产生一个随机变量 N_{n-2}，它等可能地是 $1,2,\cdots,n-2$ 中的任一个.

7. 互换 $X(N_{n-2})$ 和 $X(n-2)$ 的值.（如果 $N_2=1$，那么新排列是 $X(1)=2,X(2)=4$，$X(3)=1,X(4)=3$. 这即是最终的排列.）

8. 产生 N_{n-3}，等等. 直到产生 N_2，算法结束，交换后得到的排列即是最终的排列.

为了实现这个算法，必须产生一个随机变量，它等可能地是 $1,2,\cdots,k$ 中的任一个. 为了做到这一点，令 U 表示一个随机数，也就是说，U 服从 $(0,1)$ 上的均匀分布，并且注意到 kU 服从 $(0,k)$ 上的均匀分布. 故

$$P\{i-1<kU<i\}=\frac{1}{k} \qquad i=1,\cdots,k$$

因此，如果取 $N_k=[kU]+1$，那么 N_k 将服从期望的分布. 此处，$[x]$ 表示 x 的整数部分（也就是说，它是小于等于 x 的最大整数）.

这种算法简单地叙述如下：

步骤 1. 令 $X(1),\cdots,X(n)$ 是 $1,2,\cdots,n$ 的任一排列.（例如，令 $X(i)=i,i=1,\cdots,n$.）

步骤 2. 令 $I=n$.

步骤 3. 产生一个随机数 U，并令 $N=[IU]+1$.

步骤 4. 互换 $X(N)$ 和 $X(I)$ 的值.

步骤 5. 如果 $I>1$，将 I 的值减 1，且转到步骤 3.

步骤 6. $X(1),\cdots,X(n)$ 是要得到的随机排列.

上面的算法对产生一种随机排列是非常有用的. 例如，一个统计学家正进行一个试验，用来比较对 n 个科目的 m 种不同的处理方法的效果. 他决定将这些科目分成 m 个不同的组，每组的大小分别为 $n_1,n_2,\cdots,n_m$，其中 $\sum_{i=1}^{m} n_i = n$. 为了消除科目分配的任意偏见（例如，如果所有的"最优"科目被放在同一组中，将使试验结果失去意义），一个科目被分在一个给定的组中是"随机的"，这一点是很必要的. 如何实现呢？[1] [457]

[1] 当 $m=2$ 时，在第 6 章的例 2g 中提到了另一种方法. 这里的方法是更快的，但是需要更大的空间.

一个简单且有效的方法是对科目 1 到 n 随机编号，然后产生一种 $1,2,\cdots,n$ 的随机排列 $X(1),X(2),\cdots,X(n)$. 分配 $X(1),X(2),\cdots,X(n_1)$ 在第一组，$X(n_1+1),\cdots,X(n_1+n_2)$ 在第二组，第 j 组包含的科目编号为 $X(n_1+n_2+\cdots+n_{j-1}+k),k=1,\cdots,n_j$. ■

10.2 模拟连续型随机变量的一般方法

本节给出两种用随机数模拟连续型随机变量的一般方法.

10.2.1 逆变换法

模拟连续型随机变量的一个一般方法(称为逆变换法，inverse transformation method)是基于下面的命题的.

命题 2.1 令随机变量 U 服从(0,1)上的均匀分布. 对任意连续分布函数 F，如果用下式定义随机变量 Y：

$$Y=F^{-1}(U)$$

则随机变量 Y 的分布函数为 F. ($F^{-1}(x)=y$ 等价于 $F(y)=x$.)

证明

$$F_Y(a)=P\{Y\leqslant a\}=P\{F^{-1}(U)\leqslant a\} \tag{2-1}$$

由于 $F(x)$ 是一个单调函数，当且仅当 $U\leqslant F(a)$ 时有 $F^{-1}(U)\leqslant a$. 故由式(2-1)可以得到

$$F_Y(a)=P\{U\leqslant F(a)\}=F(a)$$

■

由命题 2.1 可知，我们可以模拟具有连续分布函数 F 的随机变量 X，方法是：首先产生一个随机变量 U，再令 $X=F^{-1}(U)$.

例 2a(模拟指数随机变量)　如果 $F(x)=1-e^{-x}$，那么 $F^{-1}(u)$ 是满足下式的 x 值：

458

$$1-e^{-x}=u$$

或者

$$x=-\log(1-u)$$

因此，如果 U 服从(0,1)上的均匀分布，那么

$$F^{-1}(U)=-\log(1-U)$$

服从均值为 1 的指数分布. 因为 $1-U$ 也服从(0,1)上的均匀分布，故 $-\log U$ 也服从均值为 1 的指数分布. 因为当 X 服从均值为 1 的指数分布时，cX 服从均值为 c 的指数分布，故 $-c\log U$ 服从均值为 c 的指数分布. ■

例 2a 的结论也可以用来模拟 Γ 随机变量.

例 2b(模拟 $\Gamma(n,\lambda)$ 随机变量)　为了模拟参数为 (n,λ) 的 Γ 分布，当 n 是整数时，利用以 λ 为参数的 n 个独立的指数随机变量之和也服从 Γ 分布这一事实. 因此，如果 $U_1,\cdots,U_n$ 是服从(0,1)上均匀分布的随机变量，则

$$X=-\sum_{i=1}^{n}\frac{1}{\lambda}\log U_i=-\frac{1}{\lambda}\log\Big(\prod_{i=1}^{n}U_i\Big)$$

服从 $\Gamma(n,\lambda)$. ■

10.2.2 拒绝法

假设有一种方法模拟具有密度函数 $g(x)$ 的随机变量. 我们以此为基础，模拟具有密度函数 $f(x)$ 的连续分布函数的随机变量，方法是由密度函数 g 模拟 Y，然后以概率 $f(Y)/g(Y)$ 接

受这个模拟值.

特别地，令 c 是一个常数，满足

$$\frac{f(y)}{g(y)} \leqslant c \quad \text{对所有的 } y$$

我们用下面的方法模拟密度函数为 f 的随机变量.

拒绝法

步骤 1. 模拟密度函数为 g 的 Y 和一个随机数 U.

步骤 2. 如果 $U \leqslant f(Y)/(cg(Y))$，令 $X=Y$. 否则返回步骤 1.

拒绝法在图 10-1 中绘图说明. 下面我们证明这一方法.

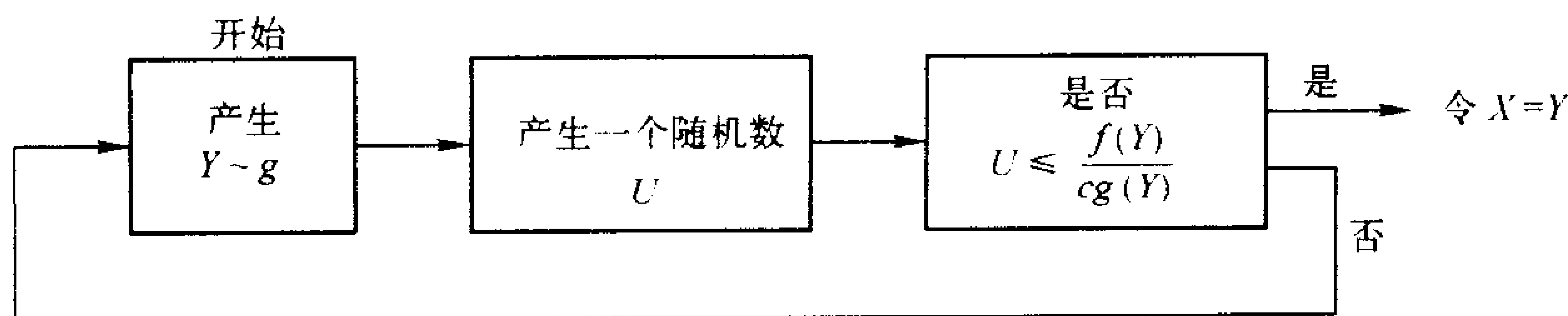


图 10-1 模拟密度函数为 f 的随机变量 X 的拒绝法

459

命题 2.2 用拒绝法产生的随机变量 X 具有密度函数 f.

证明 令 X 表示得到的值，且令 N 表示必要的迭代次数，那么

$$P\{X \leqslant x\} = P\{Y_N \leqslant x\} = P\left\{Y \leqslant x \mid U \leqslant \frac{f(Y)}{cg(Y)}\right\} = \frac{P\left\{Y \leqslant x, U \leqslant \dfrac{f(Y)}{cg(Y)}\right\}}{K}$$

其中 $K=P\{U \leqslant f(Y)/(cg(Y))\}$ 由独立性得，Y 和 U 的联合密度函数为

$$f(y,u)=g(y) \quad 0<u<1$$

因此，由上可知

$$P\{X \leqslant x\} = \frac{1}{K} \iint\limits_{\substack{y \leqslant x \\ 0 \leqslant u \leqslant f(y)/cg(y)}} g(y)\,\mathrm{d}u\,\mathrm{d}y = \frac{1}{K}\int_{-\infty}^{x}\int_{0}^{f(y)/cg(y)} \mathrm{d}u\, g(y)\,\mathrm{d}y$$

$$= \frac{1}{cK}\int_{-\infty}^{x} f(y)\,\mathrm{d}y \tag{2-2}$$

令 $x \to \infty$，并运用 f 是密度函数，得

$$1 = \frac{1}{cK}\int_{-\infty}^{\infty} f(y)\,\mathrm{d}y = \frac{1}{cK}$$

故由式(2-2)可以得到

$$P\{X \leqslant x\} = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$$

460

证毕. ■

注 (a) 应该注意，此处“以概率 $f(Y)/(cg(Y))$ 接受 Y 值”的意思是：通过产生一个随机数 U，若 $U \leqslant f(Y)/(cg(Y))$ 则接受 Y 值.

(b) 由于每次迭代都以概率 $P\{U \leqslant f(Y)/(cg(Y))\}=K=1/c$ 产生一个接受值，因此迭代次数服从均值为 c 的几何分布.

例 2c（模拟正态随机变量） 为了模拟一个标准正态随机变量 Z（也就是说，均值为 0，方差为 1)，首先注意 $|Z|$ 具有概率密度函数

$$f(x)=\frac{2}{\sqrt{2\pi}}e^{-x^2/2} \quad 0<x<\infty \tag{2-3}$$

用拒绝法模拟上述密度函数，其中 g 是服从均值为 1 的指数分布的密度函数，也就是说

$$g(x)=e^{-x} \quad 0<x<\infty$$

注意到

$$\frac{f(x)}{g(x)}=\sqrt{\frac{2}{\pi}}\exp\left\{\frac{-(x^2-2x)}{2}\right\}=\sqrt{\frac{2}{\pi}}\exp\left\{\frac{-(x^2-2x+1)}{2}+\frac{1}{2}\right\}$$

$$=\sqrt{\frac{2e}{\pi}}\exp\left\{\frac{-(x-1)^2}{2}\right\}\leqslant\sqrt{\frac{2e}{\pi}} \tag{2-4}$$

故取 $c=\sqrt{2e/\pi}$；由式(2-4)得

$$\frac{f(x)}{cg(x)}=\exp\left\{\frac{-(x-1)^2}{2}\right\}$$

因此，用拒绝法可以模拟标准正态随机变量的绝对值，方法如下：

(a) 产生独立随机变量 Y 和 U，Y 服从参数为 1 的指数分布，U 服从(0,1)上的均匀分布.

(b) 若 $U\leqslant\exp\{-(Y-1)^2/2\}$，令 $X=Y$. 否则，返回(a).

一旦模拟了具有式(2-3)中的密度函数的随机变量 X，那么可以产生一个标准正态随机变量 Z，方法是令 Z 等可能地等于 X 或 $-X$.

在步骤(b)中，若 $U\leqslant\exp\{-(Y-1)^2/2\}$，此式等价于 $-\log U\geqslant(Y-1)^2/2$，则 Y 值被接受. 然而，例 2a 说明了 $-\log U$ 服从参数为 1 的指数分布，故步骤(a)和(b)等价于：

(a′) 产生服从参数为 1 的指数分布的独立随机变量 Y_1 和 Y_2.

(b′) 若 $Y_2\geqslant(Y_1-1)^2/2$，令 $X=Y_1$. 否则，返回(a).

461

现在假定 Y_1 被接受，则知 Y_2 大于 $(Y_1-1)^2/2$. 一个比另一个大多少呢？为了回答这个问题，回忆 Y_2 服从参数为 1 的指数分布，因此 Y_2 超出 $(Y_1-1)^2/2$ 的值(即，超过 $(Y_1-1)^2/2$ 的"额外寿命")也服从参数为 1 的指数分布(由无记忆性). 也就是说，当我们接受(b′)时，不仅可以得到 X(标准正态随机变量的绝对值)，而且通过计算 $Y_2-(Y_1-1)^2/2$ 还可以产生参数为 1 的指数随机变量(与 X 独立).

因此，总结起来，可以用下面的算法产生参数为 1 的指数随机变量和独立的标准正态随机变量：

步骤 1. 产生一个参数为 1 的指数随机变量 Y_1.

步骤 2. 产生一个参数为 1 的指数随机变量 Y_2.

步骤 3. 若 $Y_2-(Y_1-1)^2/2>0$，令 $Y=Y_2-(Y_1-1)^2/2$，并转入步骤 4. 否则，转入步骤 1.

步骤 4. 产生一个随机数 U，并令

$$Z=\begin{cases} Y_1 & 若\ U\leqslant\dfrac{1}{2} \\ -Y_1 & 若\ U>\dfrac{1}{2} \end{cases}$$

由上述方法产生的随机变量 Z 和 Y 是独立的，Z 服从均值为 0、方差为 1 的正态分布，Y 服从参数为 1 的指数分布.(如果需要均值为 μ、方差为 σ^2 的正态随机变量，只需取 $\mu+\sigma Z$). ■

注　(a) 因为 $c=\sqrt{2e/\pi}\approx1.32$，步骤 2 中需要的迭代次数平均为 1.32.

(b) 如果需要产生一个服从标准正态分布的随机变量序列，那么用步骤 3 中得到的指数随机变量 Y 作为产生下一个正态随机变量的初值(步骤 1 中需要的)．因此，平均来讲，可以通过产生 1.64(＝2×1.32－1)个指数随机变量且计算 1.32 的平方，来模拟一个标准正态随机变量．

例 2d(模拟正态随机变量——极坐标法) 第 6 章中例 7b 表明，如果 X 和 Y 是独立的标准正态随机变量，那么它们的极坐标 $R=\sqrt{X^2+Y^2}$，$\theta=\tan^{-1}(Y/X)$是独立的，R^2 服从均值为 2 的指数分布，θ 服从$(0,2\pi)$上的均匀分布．因此，若 U_1 和 U_2 是随机数，那么可以(用例 2a 的结论)令

$$R=(-2\log U_1)^{1/2}$$
$$\theta=2\pi U_2$$

得到

$$X=R\cos\theta=(-2\log U_1)^{1/2}\cos(2\pi U_2)$$
$$Y=R\sin\theta=(-2\log U_1)^{1/2}\sin(2\pi U_2) \tag{2-5}$$

462 是独立的标准正态随机变量． ■

上述产生标准正态随机变量的方法称为博克斯-米勒(Box-Muller)方法．因为需要计算正弦值和余弦值，这种方法有点儿麻烦．然而，有一种方法可以避免这种困难．首先注意到，若 U 服从$(0,1)$上的均匀分布，那么 $2U$ 服从$(0,2)$上的均匀分布，$2U-1$ 服从$(-1,1)$上的均匀分布．因此，若得到随机数 U_1 和 U_2，且令

$$V_1=2U_1-1$$
$$V_2=2U_2-1$$

则(V_1,V_2)在中心为$(0,0)$、面积为 4 的正方形内服从均匀分布(参见图 10-2)．

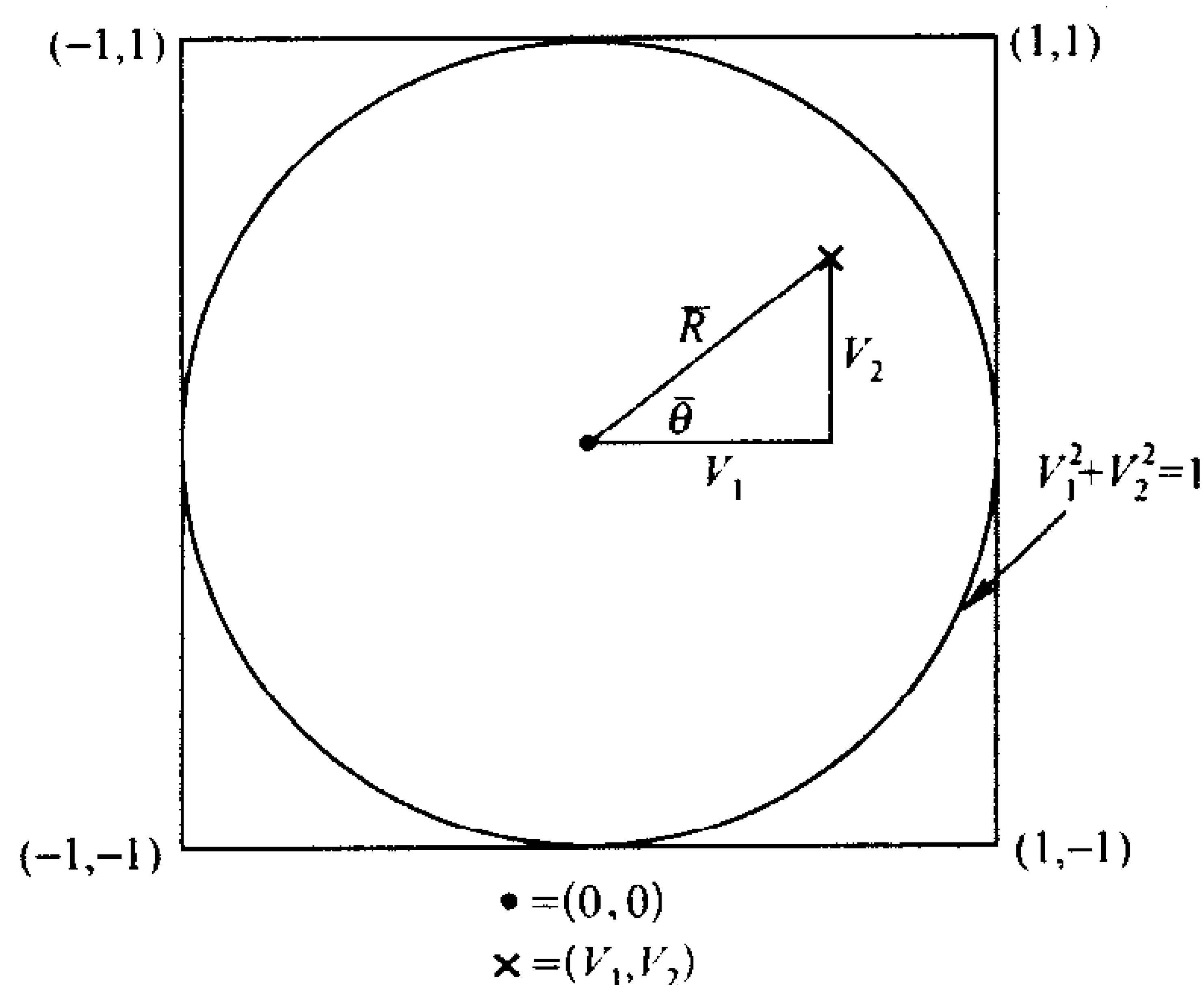


图 10-2

假设连续地产生数对(V_1,V_2)，直到得到一个在中心为$(0,0)$、半径为 1 的圆内的数对，也就是说，直到(V_1,V_2)满足 $V_1^2+V_2^2\leqslant 1$．在圆内(V_1,V_2)服从均匀分布．如果令 $\bar{R}$，$\bar{\theta}$表示数对的极坐标，那么容易证明 $\bar{R}$ 和$\bar{\theta}$是独立的，并且 $\bar{R}^2$ 服从$(0,1)$上的均匀分布，$\bar{\theta}$服从$(0,2\pi)$上的均匀分布(见习题 13)．

由于

$$\sin\bar{\theta}=\frac{V_2}{\bar{R}}=\frac{V_2}{\sqrt{V_1^2+V_2^2}}$$

$$\cos\bar{\theta}=\frac{V_1}{\bar{R}}=\frac{V_1}{\sqrt{V_1^2+V_2^2}}$$

因此由式(2-5)可知，通过产生另一个随机数 U，且令

$$X=(-2\log U)^{1/2}V_1/\bar{R}$$

$$Y=(-2\log U)^{1/2}V_2/\bar{R}$$

463

可以产生独立的标准正态随机变量 X 和 Y. 事实上，因为 $\bar{R}^2$ 服从(0,1)上的均匀分布(条件为 $V_1^2+V_2^2\leqslant 1$)，且与$\bar{\theta}$是独立的，可以用它代替产生一个新的随机数 U，故有

$$X=(-2\log\bar{R}^2)^{1/2}\frac{V_1}{\bar{R}}=\sqrt{\frac{-2\log S}{S}}V_1$$

$$Y=(-2\log\bar{R}^2)^{1/2}\frac{V_2}{\bar{R}}=\sqrt{\frac{-2\log S}{S}}V_2$$

是独立的标准正态随机数，其中

$$S=\bar{R}^2=V_1^2+V_2^2$$

总之，用下面的方法可以产生一对独立的标准正态随机数：

步骤 1. 产生随机数 U_1 和 U_2.

步骤 2. 令 $V_1=2U_1-1$，$V_2=2U_2-1$，$S=V_1^2+V_2^2$.

步骤 3. 若 $S>1$，返回步骤 1.

步骤 4. 得到独立的标准正态随机数

$$X=\sqrt{\frac{-2\log S}{S}}V_1,\ Y=\sqrt{\frac{-2\log S}{S}}V_2$$

上述方法称为极坐标法. 因为正方形内的一个随机点落在圆内的概率等于 $\pi/4$(圆的面积除以正方形的面积)，因此，平均来说，极坐标法在步骤 1 中需要 $4/\pi\approx 1.273$ 次迭代. 于是，平均来说，产生两个独立的标准正态随机数需要 2.546 个随机数、1 次对数、1 次平方根、1 次除法、4.546 次乘法.

例 2c(模拟 χ^2 随机变量)　设 $Z_i(i=1,\cdots,n)$ 是 n 个独立的标准正态随机变量，则 $\chi_n^2=Z_1^2+\cdots+Z_n^2$ 服从自由度为 n 的 χ^2 分布. 第 6 章 6.3 节说明了 $Z_1^2+Z_2^2$ 服从失效率为 $\frac{1}{2}$ 的指数分布. 因此，当 n 是偶数时，令 $n=2k$，则 χ_{2k}^2 服从参数为 $\left(k,\frac{1}{2}\right)$的 Γ 分布. 于是，$-2\log\left(\prod_{i=1}^{k}U_i\right)$ 服从自由度为 $2k$ 的 χ^2 分布. 首先模拟一个标准正态随机变量 Z，然后在前面加上 Z^2，就可以模拟自由度为 $2k+1$ 的 χ^2 分布. 即

$$\chi_{2k+1}^2=Z^2-2\log\left(\prod_{i=1}^{k}U_i\right)$$

此处 $Z,U_1,\cdots,U_n$ 相互独立，且 Z 服从标准正态分布，其他随机变量服从(0,1)上的均匀分布. ■

464

10.3 离散分布的模拟

对连续分布随机变量的模拟方法可以类似地运用到离散情形．例如，如果要模拟一个随机变量 X，其概率质量函数为

$$P\{X=x_j\}=P_j, j=0,1,\cdots, \sum_j P_j=1$$

可以用逆变换法模拟上面的离散时间．

为了模拟满足 $P\{X=x_j\}=P_j$ 的 X，令 U 服从 $(0,1)$ 上的均匀分布，并令

$$X=\begin{cases} x_1 & \text{若 } U\leqslant P_1 \\ x_2 & \text{若 } P_1<U\leqslant P_1+P_2 \\ \vdots & \\ x_j & \text{若 } \sum_1^{j-1}P_i<U\leqslant\sum_1^{j}P_i \\ \vdots & \end{cases}$$

由于

$$P\{X=x_j\}=P\left\{\sum_1^{j-1}P_i<U\leqslant\sum_1^{j}P_i\right\}=P_j$$

可以看到 X 具有所期望的分布．

例 3a（几何分布） 假设独立试验成功的概率为 p，$0<p<1$，直到有一次成功发生，试验结束．令 X 表示试验次数，则

$$P\{X=i\}=(1-p)^{i-1}p \quad i\geqslant 1$$

此处表示前 $i-1$ 次试验失败，第 i 次试验成功．随机变量 X 被称为服从参数为 p 的几何分布．因为

$$\begin{aligned}\sum_{i=1}^{j-1}P\{X=i\}&=1-P\{X>j-1\}=1-P\{\text{前 } j-1 \text{ 次均失败}\}\\&=1-(1-p)^{j-1} \quad j\geqslant 1\end{aligned}$$

可以模拟这样一个随机变量：通过产生一个随机数 U，并令 X 等于 j，满足

465

$$1-(1-p)^{j-1}<U\leqslant 1-(1-p)^j$$

上式等价于

$$(1-p)^j\leqslant 1-U<(1-p)^{j-1}$$

由于 $1-U$ 与 U 有相同的分布，可以定义 X 如下：

$$X=\min\{j:(1-p)^j\leqslant U\}=\min\{j:j\log(1-p)\leqslant\log U\}=\min\left\{j:j\geqslant\frac{\log U}{\log(1-p)}\right\}$$

此处不等式变号是由于 $\log(1-p)$ 是负的（因为 $\log(1-p)<\log 1=0$）．用记号 $[x]$ 表示 x 的整数部分（也就是说，$[x]$ 表示小于或等于 x 的整数），可以记作

$$X=1+\left[\frac{\log U}{\log(1-p)}\right]$$

■

正如连续分布情况，特殊的模拟方法已经用到离散分布情况，下面给出其中的两种．

例 3b（模拟二项随机变量） 通过表示成 n 个独立伯努利随机变量之和，很容易模拟一个服从参数为 (n,p) 的二项分布的随机变量．也就是说，如果 $U_1,\cdots,U_n$ 独立且服从 $(0,1)$ 上的均

匀分布，然后令

$$X_i = \begin{cases} 1 & 若 U_i < p \\ 0 & 其他 \end{cases}$$

则 $X \equiv \sum_{i=1}^{n} X_i$ 服从参数为 n 和 p 的二项分布. ■

例 3c（模拟泊松随机变量） 为了模拟一个均值为 λ 的泊松随机变量，产生服从(0,1)上均匀分布的独立随机变量 $U_1, U_2 \cdots$，满足下式时停止：

$$N = \min\left\{n: \prod_{i=1}^{n} U_i < e^{-\lambda}\right\}$$

随机变量 $X \equiv N-1$ 具有所期望的分布. 也就是说，如果继续产生随机数直到它们的乘积小于 $e^{-\lambda}$，那么得到的数减 1 就服从均值为 λ 的泊松分布.

$X \equiv N-1$ 确实服从均值为 λ 的泊松分布，这一点很容易从下面看到

$$X + 1 = \min\left\{n: \prod_{i=1}^{n} U_i < e^{-\lambda}\right\}$$

466

等价于

$$X = \max\left\{n: \prod_{i=1}^{n} U_i \geqslant e^{-\lambda}\right\} \quad 其中 \prod_{i=1}^{0} U_i \equiv 1$$

或者，取对数得到

$$X = \max\left\{n: \sum_{i=1}^{n} \log U_i \geqslant -\lambda\right\}$$

或者

$$X = \max\left\{n: \sum_{i=1}^{n} -\log U_i \leqslant \lambda\right\}$$

然而，$-\log U_i$ 服从失效率为 1 的指数分布，对失效率为 1 的指数分布求和，使其小于等于 λ，X 是求和的个数. 由于强度为 1 的泊松过程的连续事件间的时间独立地服从失效率为 1 的指数分布，因此强度为 1 的泊松过程到时刻 λ 停止时的事件数即为 X，于是 X 服从均值为 λ 的泊松分布. ■

10.4 减小方差的方法

令 $X_1, \cdots, X_n$ 具有给定的联合分布，假设对计算下式感兴趣：

$$\theta \equiv E[g(X_1, \cdots, X_n)]$$

此处 g 是一个特定的函数. 有时用解析法计算上式是相当困难的，在这种情况下，我们试图用模拟法估计 θ. 具体做法如下：产生与 $X_1, \cdots, X_n$ 有相同分布的 $X_1^{(1)}, \cdots, X_n^{(1)}$，令

$$Y_1 = g(X_1^{(1)}, \cdots, X_n^{(1)})$$

现在模拟第二个随机变量序列 $X_1^{(2)}, \cdots, X_n^{(2)}$（与第一个随机变量序列独立），$X_1^{(2)}, \cdots, X_n^{(2)}$ 与 $X_1, \cdots, X_n$ 有相同的分布，令

$$Y_2 = g(X_1^{(2)}, \cdots, X_n^{(2)})$$

直到产生 k（某个预定的数）个随机变量序列，由此计算出 $Y_1, \cdots, Y_k$. $Y_1, \cdots, Y_k$ 是独立同分布的随机变量，每一个都有相同的分布密度 $g(X_1, \cdots, X_n)$. 因此，如果令 $\bar{Y}$ 表示 k 个随机变量的

平均，也就是说，

$$\bar{Y}=\sum_{i=1}^{k}\frac{Y_i}{k}$$

那么

467

$$E[\bar{Y}]=\theta$$
$$E[(\bar{Y}-\theta)^2]=\mathrm{Var}(\bar{Y})$$

因此，我们用 $\bar{Y}$ 作为 θ 的估计. $\bar{Y}$ 减去 θ 的差的平方的期望即是 $\bar{Y}$ 的方差，我们期望这个值尽可能小.（在上述情况下，$\mathrm{Var}(\bar{Y})=\mathrm{Var}(Y_i)/k$，这个值通常是未知的，必须通过产生的值 $Y_1,\cdots,Y_n$，来估计.）下面给出三种减小估计方差的一般方法.

10.4.1 利用对立变量

在前述情况下，假设已经产生了 Y_1 和 Y_2，它们是具有均值 θ 的独立随机变量. 由于

$$\mathrm{Var}\left(\frac{Y_1+Y_2}{2}\right)=\frac{1}{4}[\mathrm{Var}(Y_1)+\mathrm{Var}(Y_2)+2\mathrm{Cov}(Y_1,Y_2)]=\frac{\mathrm{Var}(Y_1)}{2}+\frac{\mathrm{Cov}(Y_1,Y_2)}{2}$$

因此，如果 Y_1 和 Y_2 不独立，而只是不相关，则结果是较好的(从某种意义上说方差将减小). 假设随机变量 $X_1,\cdots,X_n$ 是独立的，每一个随机变量都用逆变换法模拟. 也就是说，用 $F_i^{-1}(U_i)$模拟 X_i，此处 U_i 是一个随机数，F_i 是 X_i 的分布函数. 于是，Y_1 可以表示为

$$Y_1=g(F_1^{-1}(U_1),\cdots,F_n^{-1}(U_n))$$

因为只要 U 是一个随机数，$1-U$ 也服从(0,1)上的均匀分布(与 U 是不相关的)，所以 Y_2 由下式定义：

$$Y_2=g(F_1^{-1}(1-U_1),\cdots,F_n^{-1}(1-U_n))$$

Y_2 和 Y_1 有相同的分布. 因此，如果 Y_1 和 Y_2 是不相关的，那么用这个均值产生 Y_2 比用一组新的随机数产生 Y_2 有更小的方差.（另外，还将减少计算量，因为不必另外产生 n 个随机数，只需用 1 减去已产生的 n 个中的每一个.）一般而言，虽然不能确定 Y_1 和 Y_2 是不相关的，但通常可以证明是这种情况. 事实上，只要 g 是单调的，就可以证明 Y_1 和 Y_2 是不相关的.

10.4.2 利用条件期望

首先回忆一下条件方差公式(参看 7.4.4 节)：

$$\mathrm{Var}(Y)=E[\mathrm{Var}(Y|Z)]+\mathrm{Var}(E[Y|Z])$$

假设对估计 $E[g(X_1,\cdots,X_n)]$感兴趣，方法是先模拟 $\boldsymbol{X}=(X_1,\cdots,X_n)$，然后计算 $Y=g(\boldsymbol{X})$. 现在，如果对某个随机变量 Z 可以计算 $E[Y|Z]$，当 $\mathrm{Var}(Y|Z)\geqslant 0$ 时，从上面的条件方差公式得

468

$$\mathrm{Var}(E[Y|Z])\leqslant\mathrm{Var}(Y)$$

又由于 $E[E[Y|Z]]=E[Y]$，可得 $E[Y|Z]$是 $E[Y]$的一个更好的估计.

例 4a(估计 π)　令 U_1 和 U_2 是随机数，并设 $V_i=2U_i-1,i=1,2$. 正如例 2d 中所说的，(V_1,V_2)在中心为(0,0)、面积为 4 的正方形内服从均匀分布，落在中心为(0,0)、半径为 1 的圆内的概率为 $\pi/4$(圆的面积与正方形面积的比值). 见图 10-2. 因此，模拟 n(较大的数)个点，令

$$I_j=\begin{cases}1 & \text{若第 } j \text{ 个点落在圆内}\\ 0 & \text{其他}\end{cases}$$

$I_j(j=1,\cdots,n)$是独立同分布的随机变量，且 $E[I_j]=\pi/4$. 于是，由强大数定律得

$$\frac{I_1+\cdots+I_n}{n}\to\frac{\pi}{4},\quad 当 n\to\infty 时$$

因此，通过模拟很多的(V_1,V_2)，它们落在圆内的比例乘以4，就可以精确地近似π.

然而，运用条件期望可以提高上述估计. 如果令I是(V_1,V_2)的示性变量，那么在条件V_1上用I的观察值是更好的，因此利用

$$E[I|V_1]=P\{V_1^2+V_2^2\leqslant 1|V_1\}=P\{V_2^2\leqslant 1-V_1^2|V_1\}$$

因为

$$P\{V_2^2\leqslant 1-V_1^2|V_1=v\}=P\{V_2^2\leqslant 1-v^2\}=P\{-\sqrt{1-v^2}\leqslant V_2\leqslant\sqrt{1-v^2}\}=\sqrt{1-v^2}$$

所以

$$E[I|V_1]=\sqrt{1-V_1^2}$$

于是，用I的平均值得到的改进实际上是用$\sqrt{1-V_1^2}$的平均值得到的改进. 因为

$$E\left[\sqrt{1-V_1^2}\right]=\int_{-1}^{1}\frac{1}{2}\sqrt{1-v^2}\mathrm{d}v=\int_0^1\sqrt{1-u^2}\mathrm{d}u=E\left[\sqrt{1-U^2}\right]$$

此处U服从(0,1)上的均匀分布，所以可产生n个随机数U，然后用$\sqrt{1-U^2}$的平均值作为$\pi/4$的估计.（习题14表明了这个估计和$\sqrt{1-V^2}$的平均值有相同的方差.）

469

注意到函数$g(u)=\sqrt{1-u^2}(0\leqslant u\leqslant 1)$关于$u$单调递减，可以进一步提高$\pi$的上述估计，因此对立变量的方法将减小$E\left[\sqrt{1-U^2}\right]$的估计的方差. 也就是说，通过产生$n/2$个随机数$U$，然后用$\sqrt{1-U^2}+\sqrt{1-(1-U)^2}$的平均值的一半估计$\pi/4$，比产生$n$个随机数$U$，用$\sqrt{1-U^2}$的平均值估计$\pi/4$要更精确.

用$n=10\,000$，基于三个估计值，下表分别给出了π的模拟结果.

方　法	π的估计值
随机点落在圆内的比例	3.161 2
$\sqrt{1-U^2}$的平均值	3.128 448
$\sqrt{1-U^2}+\sqrt{1-(1-U)^2}$的平均值	3.139 578

当$n=64\,000$时，用最后一种方法进行模拟，可以得到π的估计值为3.143 288. ■

10.4.3 控制变量

再次假设要用模拟来估计$E[g(\boldsymbol{X})]$，此处$\boldsymbol{X}=(X_1,\cdots,X_n)$. 但此处假设对某个函数$f$，$f(\boldsymbol{X})$的数学期望是已知的，即$E[f(\boldsymbol{X})]=\mu$. 那么，对任一常数$a$，也可以用

$$W=g(\boldsymbol{X})+a[f(\boldsymbol{X})-\mu]$$

作为$E[g(\boldsymbol{X})]$的一个估计. 由

$$\mathrm{Var}(W)=\mathrm{Var}[g(\boldsymbol{X})]+a^2\mathrm{Var}[f(\boldsymbol{X})]+2a\mathrm{Cov}[g(\boldsymbol{X}),f(\boldsymbol{X})]\tag{4-1}$$

通过简单的计算可以得到，当

$$a=\frac{-\mathrm{Cov}[f(\boldsymbol{X}),g(\boldsymbol{X})]}{\mathrm{Var}[f(\boldsymbol{X})]}\tag{4-2}$$

时，上式达到最小值. 对此时的 a，

$$\mathrm{Var}(W)=\mathrm{Var}[g(\boldsymbol{X})]-\frac{[\mathrm{Cov}[f(\boldsymbol{X}),g(\boldsymbol{X})]]^2}{\mathrm{Var}[f(\boldsymbol{X})]} \tag{4-3}$$

可是，无论 $\mathrm{Var}[f(\boldsymbol{X})]$还是 $\mathrm{Cov}[f(\boldsymbol{X}),g(\boldsymbol{X})]$通常都是未知的，因此通常无法得到上述方差的减小量. 实际中运用的一种方法是用模拟数据估计这些量. 通常，这种方法理论上可以减小方差.

470

小结

令 F 是一个连续分布函数，U 是服从(0,1)上均匀分布的随机变量. 随机变量 $F^{-1}(U)$的分布函数为 F，此处 $F^{-1}(U)$的值为 x，x 满足 $F(x)=u$. 应用这一结论，可以通过服从(0,1)上均匀分布的随机变量的值(称为*随机数*)产生其他的随机变量的值. 这种方法称为*逆变换法*.

产生随机数的另一种方法基于*拒绝法*. 假设有一个有效的方法产生一个密度函数为 g 的随机变量，想要产生密度函数为 f 的随机变量. 通过确定一个常数 c，利用拒绝法可以实现这一点，常数 c 满足

$$\max\frac{f(x)}{g(x)}\leqslant c$$

然后如下继续：

1. 产生密度函数为 g 的随机变量 Y.
2. 产生一个随机数 U.
3. 若 $U\leqslant f(Y)/(cg(Y))$，令 $X=Y$，并且停止.
4. 转回步骤 1.

步骤 1 产生的随机变量服从均值为 c 的几何分布.

用拒绝法或者极坐标算法可以有效地模拟标准正态随机变量(用均值为 1 的指数分布 g).

为了估计 θ，通常产生期望值为 θ 的部分随机数. 当这些随机变量有较小的方差时，可以增加这种方法的有效性. 用来产生均值为 θ 的随机变量，下面的三种方法相对来说方差较小：

1. 运用对立变量.
2. 运用条件期望.
3. 运用控制变量.

习题

1. 下面的算法将产生元素 $1,2,\cdots,n$ 的随机排列. 这种方法有时比例 1a 更快，但元素的位置不固定，直到算法结束. 在这个算法中，认为位置 i 上的元素为 $P(i)$.

   步骤 1. 令 $k=1$.

   步骤 2. 令 $P(1)=1$.

471

   步骤 3. 若 $k=n$，终止. 否则，令 $k=k+1$.

   步骤 4. 产生一个随机数 U，并令

$$P(k)=P([kU]+1)$$
$$P([kU]+1)=k$$

转回步骤 3.

(a) 简单解释一下这个算法做的是什么.

(b) 证明：迭代 k 步，当 $P(k)$ 在最前面时，$P(1),P(2),\cdots,P(k)$ 是 $1,2,\cdots,k$ 的一个随机排列.

提示：用归纳法证明

$$P_k\{i_1,i_2,\cdots,i_{j-1},k,i_j,\cdots,i_{k-2},i\}=P_{k-1}\{i_1,i_2,\cdots,i_{j-1},i,i_j,\cdots,i_{k-2}\}\frac{1}{k}$$

$$=\frac{1}{k!}\quad\text{（运用归纳假设）}$$

2. 找出一种方法来模拟一个随机变量，其密度函数为

$$f(x)=\begin{cases}e^{2x} & -\infty<x<0\\ e^{-2x} & 0<x<\infty\end{cases}$$

3. 给出一种方法来模拟具有下列概率密度函数的随机变量：

$$f(x)=\begin{cases}\frac{1}{2}(x-2) & 2\leqslant x\leqslant 3\\ \frac{1}{2}\left(2-\frac{x}{3}\right) & 3<x\leqslant 6\\ 0 & \text{其他}\end{cases}$$

4. 给出一种方法来模拟一个随机变量，其分布函数为

$$F(x)=\begin{cases}0 & x\leqslant -3\\ \frac{1}{2}+\frac{x}{6} & -3<x<0\\ \frac{1}{2}+\frac{x^2}{32} & 0<x\leqslant 4\\ 1 & x>4\end{cases}$$

5. 用逆变换法给出一种方法，用来产生一个服从韦布尔分布的随机变量，分布函数为

$$F(t)=1-e^{-\alpha t^{\beta}}\qquad t\geqslant 0$$

6. 给出一种方法来模拟具有下列失效率函数的随机变量：

(a) $\lambda(t)=c$；　　(b) $\lambda(t)=ct$；

(c) $\lambda(t)=ct^2$；　　(d) $\lambda(t)=ct^3$.

7. 在下面的问题中，F 是分布函数

$$F(x)=x^n\qquad 0<x<1$$

472

(a) 给出一种方法，仅用一个随机数模拟具有分布函数 F 的随机变量.

(b) 令 $U_1,\cdots,U_n$ 是独立的随机数. 证明

$$P\{\max(U_1,\cdots,U_n)\leqslant x\}=x^n$$

(c) 用(b)的结论给出第二种方法，用来模拟具有分布函数 F 的随机变量.

8. 假设模拟服从 $F_i(i=1,2,\cdots,n)$ 的随机变量是相对容易的，如何模拟服从下列分布的随机变量?

(a) $F(x)=\prod\limits_{i=1}^{n}F_i(x)$；　　(b) $F(x)=1-\prod\limits_{i=1}^{n}[1-F_i(x)]$.

9. 假设有一种方法来模拟具有分布函数 F_1 和 F_2 的随机变量. 解释一下如何模拟具有下列分布函数的随机变量：

$$F(x)=pF_1(x)+(1-p)F_2(x)\quad 0<p<1$$

给出一种方法模拟具有下列分布函数的随机变量：

$$F(x)=\begin{cases}\frac{1}{3}(1-e^{-3x})+\frac{2}{3}x & 0<x\leqslant 1\\ \frac{1}{3}(1-e^{-3x})+\frac{2}{3} & x>1\end{cases}$$

10. 在例 2c 中，我们基于失效率为 1 的指数分布，用拒绝法模拟了一个标准正态随机变量的绝对值. 运用不同的指数分布密度函数，是否可以得到一种更有效的算法？此处指数分布的密度函数为 $g(x)=\lambda e^{-\lambda x}$. 试证明：当 $\lambda=1$ 时，在拒绝法中需要迭代的平均次数最小.
11. 用拒绝法，其中 $g(x)=1(0<x<1)$，确定一种算法，用来模拟具有下列密度函数的随机变量：

$$f(x)=\begin{cases}60x^3(1-x)^2 & 0<x<1\\ 0 & \text{其他}\end{cases}$$

12. 说明如何用随机数近似 $\int_0^1 k(x)dx$，此处 $k(x)$ 是任意函数.

   提示：若 U 服从(0,1)上的均匀分布，$E[k(U)]$表示什么？
13. 令(X,Y)在中心为原点、半径为 1 的圆内服从均匀分布. 因此它的联合密度为

$$f(x,y)=\frac{1}{\pi} \qquad 0\leqslant x^2+y^2\leqslant 1$$

   令 $R=(X^2+Y^2)^{1/2},\theta=\tan^{-1}(Y/X)$表示对应的极坐标. 试证明：$R$ 和 θ 是独立的，并且 R^2 服从(0,1)上的均匀分布，θ 服从$(0,2\pi)$上的均匀分布.
14. 在例 4a 中证明了

$$E[(1-V^2)^{1/2}]=E[(1-U^2)^{1/2}]=\frac{\pi}{4}$$

   当 V 服从(−1,1)上的均匀分布，U 服从(0,1)上的均匀分布时，试证明：

$$\mathrm{Var}[(1-V^2)^{1/2}]=\mathrm{Var}[(1-U^2)^{1/2}]$$

   473 并求出上式的值.
15. (a) 验证当 a 由式(4-2)给定时，式(4-1)成立的最小值.

   (b) 验证式(4-1)由式(4-3)给出时的最小值.
16. 令 X 是(0,1)上的随机变量，它的密度函数为 $f(x)$. 试证明可以通过模拟 X 来估计 $\int_0^1 g(x)dx$，然后取 $g(X)/f(X)$ 作为估计. 这种方法称为重要性抽样法，尽量选择 f 在形状上与 g 相似，以使得 $g(X)/f(X)$ 有较小的方差.

自测题与练习

1. 随机变量 X 具有概率密度函数

$$f(x)=Ce^x \qquad 0<x<1$$

   (a) 求出常数 C 的值.

   (b) 给出一种方法模拟这样一个随机变量.
2. 给出一种方法模拟一个具有下列概率密度函数的随机变量：

$$f(x)=30(x^2-2x^3+x^4) \qquad 0<x<1$$

3. 给出一种有效的算法来模拟一个随机变量的值，其概率质量函数为

$$p_1=0.15 \quad p_2=0.2 \quad p_3=0.35 \quad p_4=0.30$$

4. 如果 X 是均值为 μ、方差为 σ^2 的正态随机变量，定义一个随机变量 Y，与 X 具有相同的分布，且与 X 不相关.
5. 令 X 和 Y 是均值为 1 的独立指数随机变量.

(a) 说明如何运用模拟来估计 $E[e^{XY}]$.

(b) 用控制变量说明如何提高(a)中的估计方法.

参考文献

ROSS, S. M. *Simulation*. San Diego, Calif.: Academic Press, Inc., 1997.

474

附录A　部分习题参考答案

第1章

1. 67 600 000;19 656 000　**2.** 1 296　**4.** 24;4　**5.** 144;18　**6.** 2 401　**7.** 720;72;144;72　**8.** 120;1 260;34 650　**9.** 27 720　**10.** 40 320;10 080;1 152;2 880;384　**11.** 720;72;144　**12.** 24 300 000;17 100 720　**13.** 190　**14.** 2 598 960　**16.** 42;94　**17.** 604 800　**18.** 600　**19.** 896;1 000;910　**20.** 36;26　**21.** 35　**22.** 18　**23.** 48　**25.** $52!/(13!)^4$　**27.** 27 720　**28.** 65 536; 2 520　**29.** 12 600;945　**30.** 564 480　**31.** 165; 35　**32.** 1 287; 14 112　**33.** 220;572

第2章

9. 74　**10.** 0.4;0.1　**11.** 70;2　**12.** 0.5;0.32;149/198　**13.** 20 000;12 000;11 000;68 000;10 000　**14.** 1.057　**15.** 0.002 0;0.422 6;0.047 5;0.021 1;0.000 24　**17.** $9.109\,47\times10^{-6}$　**18.** 0.048　**19.** 5/18　**20.** 0.905 2　**22.** $(n+1)/2^n$　**23.** 5/12　**25.** 0.4　**26.** 0.492 929　**27.** 0.583 33　**28.** 0.088 8;0.247 7;0.124 3;0.209 9　**30.** 1/18;1/6;1/2　**31.** 2/9;1/9　**33.** 70/323　**35.** 0.836 3　**36.** 0.004 5;0.058 8　**37.** 0.083 3;0.5　**38.** 4　**39.** 0.48　**40.** 1/64;21/64;36/64;6/64　**41.** 0.517 7　**44.** 0.3;0.2;0.1　**46.** 5　**48.** $1.060\,4\times10^{-3}$　**49.** 0.432 9　**50.** $2.608\,4\times10^{-6}$　**52.** 0.091 45;0.426 8　**53.** 12/35　**54.** 0.051 1　**55.** 0.219 8;0.034 3

第3章

1. 1/3　**2.** 1/6;1/5;1/4;1/3;1/2;1　**3.** 0.339　**5.** 6/91　**6.** 1/2　**7.** 2/3　**8.** 1/2　**9.** 7/11　**10.** 0.22　**11.** 0.484 8　**12.** 0.983 5　**13.** 0.079 2;0.264　**14.** 0.331;0.383;0.286;48.62　**15.** 44.29;41.18　**16.** 0.4;1/26　**17.** 0.496;3/14;9/62　**18.** 0.504;0.362 9　**20.** 35/768;210/768　**21.** 4/9;1/2　**22.** 1/3;1/2　**24.** 20/21;40/41　**26.** 3/128;29/1 536　**27.** 0.089 3　**28.** 7/12;3/5　**29.** 5/11　**30.** 27/31　**31.** 3/4　**32.** 1/2　**33.** 1/3;1/5;1　**34.** 12/37　**35.** 46/185　**36.** 3/13;5/13;5/52;15/52　**37.** 43/459　**38.** 34.48　**39.** 4/9　**41.** 1/11　**44.** 2/3　**45.** 19/268　**46.** 17.5;38/165;17/33　**47.** 0.65;56/65;8/65;1/65;14/35;12/35;9/35　**48.** 0.11;16/89;12/27;3/5;9/25　**51.** 9　**53.** (c) 2/3　**56.** 2/3;1/3;3/4　**57.** 1/6;3/20　**61.** 9/13;1/2　**65.** 9;9;18;110;4;4;8;120 all over 128　**66.** 1/9;1/18　**67.** 38/64;13/64;13/64　**69.** 1/16;1/32;5/16;1/4;31/32　**70.** 9/19　**71.** 3/4;7/12　**73.** $p^2/(1-2p+2p^2)$　**74.** 0.555 0　**76.** 0.953 0　**78.** 0.5;0.6;0.8　**79.** 9/19;6/19;4/19;7/15;53/165;7/33　**84.** 97/142;15/26;33/102

第4章

1. $p(4)=6/91;p(2)=8/91;p(1)=32/91;p(0)=1/91;p(-1)=16/91;p(-2)=28/91$　**4.** 1/2;5/18;5/36;5/84;5/252;1/252;0;0;0;0　**5.** $n-2i;i=0,\cdots,n$　**6.** $p(3)=p(-3)=1/8;p(1)=p(-1)=3/8$　**12.** $p(4)=1/16;p(3)=1/8;p(2)=1/16;p(0)=1/2;p(-i)=p(i);p(0)=1$　**13.** $p(0)=0.28;p(500)=0.27,p(1\,000)=0.315;p(1\,500)=0.09;p(2\,000)=0.045$　**14.** $p(0)=1/2;p(1)=1/6;p(2)=1/12;p(3)=1/20;p(4)=1/5$　**17.** 1/4;1/6;1/12;1/2　**19.** 1/2;1/10;1/5;1/10;1/10　**20.** 0.591 8;no;−0.108　**21.** 39.28;37　**24.** $p=11/18$,maximum=23/72　**26.** 11/2;17/5　**27.** $A(p+1/10)$　**28.** 3/5　**31.** p^*　**32.** $11-10(0.9)^{10}$　**33.** 3　**35.** −0.067;1.089　**37.** 82.2;84.5　**39.** 3/8　**40.** 11/243　**42.** $p\geqslant1/2$　**45.** 3　**50.** 1/10;1/10　**51.** $e^{-0.2};1-1.2e^{-0.2}$　**53.** $1-e^{-0.6};1-e^{-219.18}$　**56.** 253　**57.** 0.576 8;0.607 0　**59.** 0.393 5;0.303 3;0.090 2　**60.** 0.888 6　**61.** 0.408 2　**63.** 0.082 1;0.242 4　**65.** 0.393 5;0.229 3;0.393 5　**66.** 0.150 0;0.101 2　**68.** 5.812 5　**69.** 32/243;486 4/6 561;160/729;160/729　**73.** $18(17)^{n-1}/(35)^n$　**76.** 3/10;5/6;75/138　**77.** 0.343 9

第5章

2. $3.5e^{-5/2}$　**3.** no;no　**4.** 1/2　**5.** $1-(0.01)^{1/5}$　**6.** $4,0,\infty$　**7.** 3/5;6/5　**8.** 2　**10.** 2/3;2/3　**11.** 2/5　**13.** 2/3;1/3　**15.** 0.797 7;0.682 7;0.369 5;0.952 2;0.158 7　**16.** $(0.993\,8)^{10}$　**18.** 22.66　**19.** 14.56

20. 0.999 4;0.75;0.977 **22.** 9.5;0.001 9 **23.** 0.925 8;0.176 2 **26.** 0.060 6;0.052 5 **28.** 0.836 3 **29.** 0.999 3 **32.** e^{-1};$e^{-1/2}$ **34.** e^{-1};1/3 **38.** 3/5 **40.** $1/y$

第 6 章

2. (a) 14/39;10/39;10/39;5/39 (b) 84;70;70;70;40;40;40;429 除以 15 **3.** 15/26;5/26;5/26;1/26 **4.** 25/169;40/169;40/169;64/169 **7.** $p(i,j)=p^2(1-p)^{i+j}$ **8.** $c=1/8$;$E[X]=0$ **9.** $(12x^2+6x)/7$;15/56;0.862 5;5/7;8/7 **10.** 1/2;$1-e^{-a}$ **11.** 0.145 8 **12.** $39.3e^{-5}$ **13.** 1/6;1/2 **15.** $\pi/4$ **16.** $n(1/2)^{n-1}$ **17.** 1/3 **18.** 7/9 **19.** 1/2 **21.** 2/5;2/5 **22.** no;1/3 **23.** 1/2;2/3;1/20;1/18 **25.** $e^{-1}/i!$ **30.** e^{-2};$1-3e^{-2}$ **32.** 0.032 6 **33.** 0.337 2;0.206 1 **34.** 0.082 9;0.376 6 **35.** 1/3;2/3;5/12;7/12 **36.** 5/13;8/13 **37.** 1/6;5/6;1/4;3/4 **42.** $(y+1)^2xe^{-x(y+1)}$;xe^{-xy};e^{-x} **43.** $1/2+3y/(4x)-y^3/(4x^3)$ **47.** $(1-2d/L)^3$ **48.** 0.792 97 **49.** $1-e^{-5\lambda a}$;$(1-e^{-\lambda a})^5$ **51.** r/π **52.** r **55.** (a)$u/(v+1)^2$

第 7 章

1. 52.5/12 **2.** 324;198.8 **5.** 3/2 **6.** 35 **7.** 0.9;4.9;4.2 **8.** $(1-(1-p)^N)/p$ **10.** 0.6;0 **11.** $2(n-1)p(1-p)$ **12.** $(3n^2-n)/(4n-2)$,$3n^2/(4n-2)$ **14.** $m/(1-p)$ **15.** 1/2 **18.** 4 **21.** 0.930 1;87.575 5 **22.** 14.7 **23.** 147/110 **26.** $n/(n+1)$;$1/(n+1)$ **29.** $\frac{437}{35}$;12;4;$\frac{123}{35}$ **31.** 175/6 **33.** 14 **34.** 20/19;360/361 **35.** 21.2;18.929;49.214 **36.** $-n/36$ **37.** 0 **38.** 1/8 **41.** 6;112/33 **42.** 100/19;16 200/6 137;10/19;3 240/6 137 **45.** 1/2;0 **47.** $1/(n-1)$ **48.** 6;7;5.819 2 **49.** 9/5;6/5;3/5;0 **50.** $2y^2$ **51.** $y^3/4$ **53.** 12 **54.** 8 **56.** $N(1-e^{-10/N})$ **57.** 12.5 **63.** −96/145 **65.** 5.16 **66.** 218 **67.** $x[1+(2p-1)^2]^n$ **69.** 1/2;1/16;2/81 **70.** 1/2 1/3 **72.** $1/i$;$[i(i+1)]^{-1}$;∞ **73.** μ;$1+\sigma^2$;是;σ^2

第 8 章

1. $\geqslant$19/20 **2.** 15/17;$\geqslant$3/4;$\geqslant$10 **3.** $\geqslant$3 **4.** $\leqslant$4/3;0.842 8 **5.** 0.141 6 **6.** 0.943 1 **7.** 0.308 5 **8.** 0.693 2 **9.** 66.564 **10.** 117 **11.** $\geqslant$0.057 **13.** 0.016 2;0.000 3;0.251 4;0.251 4 **14.** $n\geqslant 23$ **17.** $\leqslant$0.2 **22.** 0.769;0.357;0.426 7;0.109 3;0.112 184

第 9 章

1. 1/9;5/9 **3.** 0.973 5;0.909 8;0.735 8;0.557 8 **10.** (b) 1/6 **14.** 2.585;0.541 7;3.126 7 **15.** 5.509 8

附录B 自测题与练习参考答案

第1章

1. (a) 字母C, D, E, F有4! 种不同的排序．对这些排序中的每一种，在5个位置中的任一个位置插入AB或BA，可以得到一种A和B相邻的排序．也就是说，或者在C，D，E，F排列的第一个字母之前，或者在第一个和第二个字母之间，等等．因此有$2\cdot 5\cdot 4!=240$种排列．解这个问题的另一种方法是：假设AB看成一个元素，A在B之前有5! 种排序．同样B在A之前也有5! 种排序，我们再次得到共有$2\cdot 5!=240$种不同的排列．

(b) 共有$6!=720$种排列，因为A在B之前和B在A之前有相同的排列数，都有360种排列．

(c) 在720种可能的排列中，A在B之前且B在C之前的排列数与A，B，C间3! 种排序中的任一种的排列数是一样的，因此有$720/6=120$种可能的排序．

(d) A在B之前的360种排列中，有一半是C在D之前，有一半是D在C之前，因此，A在B之前且C在D之前的排列数有180种．

(e) AB看成一个元素，CD看成一个元素，且A在B之前，C在D之前，这样产生$4!=24$种不同的排序．因为A和B以及C和D的位置可交换，因此有$4\cdot 24=96$种不同的排列．

(f) E在最后有5! 种排列．因此，E不在最后有$6!-5!=600$种不同的排序．

2. $3!\ 4!\ 3!\ 3!$种坐法，因为3个国家有3! 种排序，然后每个国家的人要再排序．

3. (a) $10\cdot 9\cdot 8=720$

(b) $8\cdot 7\cdot 6+2\cdot 3\cdot 8\cdot 7=672$

得到上面的结果是因为：不包括A或B有$8\cdot 7\cdot 6$种选择，包括A和B中的一个而不包括另一个有$3\cdot 8\cdot 7$种选择．后者是因为A和B可以被安排到3个办公室中的任一个，下一个位置可以由其他的8个人之一填满，最后的位置可以安排剩余的7个人中的任一个．

(c) $8\cdot 7\cdot 6+3\cdot 2\cdot 8=384$　　(d) $3\cdot 9\cdot 8=216$　　(e) $9\cdot 8\cdot 7+9\cdot 8=576$

4. (a) $\binom{10}{7}$　　(b) $\binom{5}{3}\binom{5}{4}+\binom{5}{4}\binom{5}{3}+\binom{5}{5}\binom{5}{2}$

5. $\binom{7}{3,2,2}=210$

6. 三个位置放字母有$\binom{7}{3}=35$种选择．对每一种选择，有$(26)^3(10)^4$种不同的牌照．因此，共有$35\cdot(26)^3(10)^4$种不同的牌照．

7. n项中r项的任意选择等于$n-r$项的一个选择，即那些不被选择的项．

8. (a) $10\cdot 9\cdot 9\cdots 9=10\cdot 9^{n-1}$

(b) $\binom{n}{i}9^{n-i}$．因为第i个位置放0有$\binom{n}{i}$种选择，其他$n-1$个位置中的任一个可以放$1,\cdots,9$中的任一个数．

9. (a) $\binom{3n}{3}$　　(b) $3\binom{n}{3}$　　(c) $\binom{3}{1}\binom{2}{1}\binom{n}{2}\binom{n}{1}=3n^2(n-1)$

(d) n^3　　(e) $\binom{3n}{3}=3\binom{n}{3}+3n^2(n-1)+n^3$

10. 没有数字重复时，有$9\cdot 8\cdot 7\cdot 6\cdot 5$个数．仅有一个指定的数字出现两次时，有$\binom{5}{2}\cdot 8\cdot 7\cdot 6$个数，因此有一个数字出现两次时有$9\binom{5}{2}\cdot 8\cdot 7\cdot 6$个数．有两个指定的数字出现两次时，有$7\cdot \frac{5!}{2!\ 2!}$个数，因

此两个数字出现两次共有$\binom{9}{2}7\cdot\frac{5!}{2!\ 2!}$个数．于是答案是

$$9\cdot8\cdot7\cdot6\cdot5+9\binom{5}{2}\cdot8\cdot7\cdot6+\binom{9}{2}7\cdot\frac{5!}{2!\ 2!}$$

11. 有$\frac{(2n)!}{2^n}$种方法将参赛队员分为第一队，第二队，等等．因此，当这些队的排序不相关时，第一轮有$\frac{(2n)!}{n!\ 2^n}$种不同的安排．在第一轮的每场比赛中，对 n 个获胜者有 2^n 种可能的结果，因此，第一轮比赛中共有$\frac{(2n)!}{n!}$种不同的结果．

12. $\binom{8}{3}\binom{7}{3}+\binom{8}{4}\binom{7}{2}=3\ 430$

第一项表明委员会包括 3 位女士和 3 位男士；第二项说明委员会包括 4 位女士和 2 位男士.

13. ($x_1+\cdots+x_5=4$ 的解决方案数)($x_1+\cdots+x_5=5$ 的解决方案数)($x_1+\cdots+x_5=6$ 的解决方案数)$=\binom{8}{4}\binom{9}{4}\binom{10}{4}$.

14. 因为有$\binom{j-1}{n-1}$个正向量，总数是 j，故满足题意的向量有 $\sum_{j=n}^{k}\binom{j-1}{n-1}$ 个．

15. 首先确定 k 个人通过考试的不同的结果数．因为对 k 个人有$\binom{n}{k}$种不同的分组，且他们的成绩有 $k!$ 种不同的排序，因此 k 个人通过考试有$\binom{n}{k}k!$ 种可能的结果．于是，共有 $\sum_{k=0}^{n}\binom{n}{k}k!$ 种可能的结果．

第 2 章

1. (a) $2\cdot3\cdot4=24$　(b) $2\cdot3=6$

(c) $3\cdot4=12$　(d) $AB=\{(c$ 意大利面条$,i),(c,$米饭$,i),(c,$土豆$,i)\}$

(e) 8　(f) $ABC=\{(c,$米饭$,i)\}$

2. 令 A 表示购买套装，B 表示购买衬衣，C 表示购买领带，则

$$P(A\cup B\cup C)=0.22+0.30+0.28-0.11-0.14-0.10+0.06=0.51$$

(a) $1-0.51=0.49$

(b) 购买两种或更多种物品的概率为

$$P(AB\cup AC\cup BC)=0.11+0.14+0.10-0.06-0.06-0.06+0.06=0.23$$

故只购买一种物品的概率为 $0.51-0.23=0.28$.

3. 由对称性知，第 14 张牌等可能地是 52 张牌中的任一张，因此概率是 4/52. 一种更正式的理由是：从 52! 种结果中数一下第 14 张是 A 的次数，得到

$$p=\frac{4\cdot51\cdot50\cdots2\cdot1}{(52)!}=\frac{4}{52}$$

令 B 表示第 14 张牌首次出现 A，有

$$P(B)=\frac{48\cdot47\cdots36\cdot4}{52\cdot51\cdots40\cdot39}=0.031\ 2$$

4. 令 D 表示最小温度为 70℉的事件，则

$$P(A\cup B)=P(A)+P(B)-P(AB)=0.7-P(AB)$$
$$P(C\cup D)=P(C)+P(D)-P(CD)=0.2+P(D)-P(CD)$$

上面的等式减去下面的等式，并注意到 $A\cup B=C\cup D$ 和 $AB=CD$，得

$$0=0.5-P(D)$$

也就是说，$P(D)=0.5$.

5. (a) $\frac{52 \cdot 48 \cdot 44 \cdot 40}{52 \cdot 51 \cdot 50 \cdot 49}=0.6761$ (b) $\frac{52 \cdot 39 \cdot 26 \cdot 13}{52 \cdot 51 \cdot 50 \cdot 49}=0.1055$

6. 令 R 表示两个球都是红的，B 表示两个球都是黑的，则

$$P(R \cup B)=P(R)+P(B)=\frac{3 \cdot 4}{6 \cdot 10}+\frac{3 \cdot 6}{6 \cdot 10}=1/2$$

7. (a) $\frac{1}{\binom{40}{8}}=1.3 \times 10^{-8}$ (b) $\frac{\binom{8}{7}\binom{32}{1}}{\binom{40}{8}}=3.3 \times 10^{-6}$

(c) $\frac{\binom{8}{6}\binom{32}{2}}{\binom{40}{8}}+1.3 \times 10^{-8}+3.3 \times 10^{-6}=1.8 \times 10^{-4}$

8. (a) $\frac{3 \cdot 4 \cdot 4 \cdot 3}{\binom{14}{4}}=0.1439$ (b) $\frac{\binom{4}{2}\binom{4}{2}}{\binom{14}{4}}=0.0360$ (c) $\frac{\binom{8}{4}}{\binom{14}{4}}=0.0699$

9. 令 $S=\bigcup_{i=1}^{n} A_i$，考虑从 S 中随机选择一个元素，则 $P(A)=N(A)/N(S)$，这个结果满足命题 4.3 和命题 4.4.

10. 因为有 5! =120 种结果，此处 1 号马的位置是指定的，满足 $N(A)=360$. 同样地，$N(B)=120$，$N(AB)=2 \cdot 4!=48$. 因此，由自测题 9 可以得到 $N(A \cup B)=432$.

11. 解这个问题的一种方法是：求余事件的概率，即至少有一个花色不出现的概率．令 A_i 表示第 i 个花色不出现，其中 $i=1,2,3,4$. 于是有

$$P\left(\bigcup_{i=1}^{4} A_i\right)=\sum_i P(A_i)-\sum_j \sum_{i:i<j} P(A_iA_j)+\cdots-P(A_1A_2A_3A_4)$$

$$=4\frac{\binom{39}{5}}{\binom{52}{5}}-\binom{4}{2}\frac{\binom{26}{5}}{\binom{52}{5}}+\binom{4}{3}\frac{\binom{13}{5}}{\binom{52}{5}}=4\frac{\binom{39}{5}}{\binom{52}{5}}-6\frac{\binom{26}{5}}{\binom{52}{5}}+4\frac{\binom{13}{5}}{\binom{52}{5}}$$

所求概率等于 1 减上式．另一种方法是令 A 表示四个花色都被给出，则用

$$P(A)=P(n,n,n,n,o)+P(n,n,n,o,n)+P(n,n,o,n,n)+P(n,o,n,n,n)$$

例如，此处 $P(n,n,n,o,n)$ 是这样的概率：第一张牌是新花色，第二张牌是新花色，第三张牌是新花色，第四张牌是旧花色（也就是说，这种花色已经出现过），第五张牌是新花色．

得到

$$P(A)=\frac{52 \cdot 39 \cdot 26 \cdot 13 \cdot 48+52 \cdot 39 \cdot 26 \cdot 36 \cdot 13}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}+\frac{52 \cdot 39 \cdot 24 \cdot 26 \cdot 13+52 \cdot 12 \cdot 39 \cdot 26 \cdot 13}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}$$

$$=\frac{52 \cdot 39 \cdot 26 \cdot 13(48+36+24+12)}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}=0.2637$$

12. 将 10 个队员分成第一个宿舍对，第二个宿舍对，等等有 $(10)!/2^5$ 种不同的划分．于是，分成 5 个宿舍对有 $(10)!/(5!\ 2^5)$ 种划分．在混合的宿舍对中选择前锋队员和后卫队员，有 $\binom{6}{2}\binom{4}{2}$ 种方法，然后又两种方法将他们配成对．将剩下的两个后卫队员配对有一种方法，将剩下的四个后卫队员分成两对有 $4!/(2!\ 2^2)=3$ 种方法，故所求的概率为

$$P\{2\text{ 个混合对}\}=\frac{\binom{6}{2}\binom{4}{2}(2)(3)}{(10)!/(5!\ 2^5)}=0.5714$$

13. 令 R 表示字母 R 重复；同样地，定义事件 E 和 V. 于是

$$P\{\text{字母相同}\}=P(R)+P(E)+P(V)=\frac{2}{7}\frac{1}{8}+\frac{3}{7}\frac{1}{8}+\frac{1}{7}\frac{1}{8}=\frac{3}{28}$$

14. 令 $B_1=A_1, B_i=A_i\left(\bigcup_{j=1}^{i-1}A_j\right)^c, i>1$, 则

$$P\left(\bigcup_{i=1}^{\infty}A_i\right)=P\left(\bigcup_{i=1}^{\infty}B_i\right)=\sum_{i=1}^{\infty}P(B_i)\leqslant\sum_{i=1}^{\infty}P(A_i)$$

此处最后的等号是因为 B_i 是互不相容的，不等号是因为 $B_i\subset A_i$.

15. $P\left(\bigcap_{i=1}^{\infty}A_i\right)=1-P\left(\left(\bigcap_{i=1}^{\infty}A_i\right)^c\right)=1-P\left(\bigcap_{i=1}^{\infty}A_i^c\right)\geqslant 1-\sum_{i=1}^{\infty}P(A_i^c)=1$

16. $\{1\}$是一个子集的分割数等于剩下的 $n-1$ 个元素分成 $k-1$ 个非空子集的分割数，即 $T_{k-1}(n-1)$. $\{2,\cdots,n-1\}$分成 k 个非空子集有 $T_k(n-1)$种分割，然后从 k 个集合中选择一个放入元素 1，有 $kT_k(n-1)$种分割，此处$\{1\}$不是一个子集. 因此，结论成立.

17. 令 R,W,B 分别表示：选出的球没有红球、没有白球、没有蓝球的事件，则

$$\begin{aligned}P(R\cup W\cup B)&=P(R)+P(W)+P(B)-P(RW)-P(RB)-P(WB)+P(RWB)\\&=\frac{\binom{13}{5}}{\binom{18}{5}}+\frac{\binom{12}{5}}{\binom{18}{5}}+\frac{\binom{11}{5}}{\binom{18}{5}}-\frac{\binom{7}{5}}{\binom{18}{5}}-\frac{\binom{6}{5}}{\binom{18}{5}}-\frac{\binom{5}{5}}{\binom{18}{5}}\approx 0.2933\end{aligned}$$

因此，在选出的子集中所有颜色都出现的概率近似为 $1-0.2933=0.7067$.

第 3 章

1. (a) $P(\text{没有 A})=\binom{35}{13}\Big/\binom{39}{13}$; (b) $1-P(\text{没有 A})-\frac{4\binom{35}{12}}{\binom{39}{13}}$; (c) $P(i\text{ 张 A})=\frac{\binom{3}{i}\binom{36}{13-i}}{\binom{39}{13}}$.

2. 令 L_i 表示电池的寿命比 $10\,000\times i$ 英里更长的事件.

(a) $P(L_2|L_1)=P(L_1L_2)/P(L_1)=P(L_2)/P(L_1)=1/2$

(b) $P(L_3|L_1)=P(L_1L_3)/P(L_1)=P(L_3)/P(L_1)=1/8$

3. 在一个箱子中放 1 个白球和 0 个黑球，剩下的 9 个白球和 10 个黑球放在第二个箱子中.

4. 令 T 表示转移的球为白球的事件，令 W 表示从 B 箱中取出一个白球的事件，则

$$P(T|W)=\frac{P(W|T)P(T)}{P(W|T)P(T)+P(W|T^c)P(T^c)}=\frac{(2/7)(2/3)}{(2/7)(2/3)+(1/7)(1/3)}=4/5$$

5. 令 B_i 表示第 i 个球为黑球的事件，令 $R_i=B_i^c$.

$$\begin{aligned}P(B_1|R_2)&=\frac{P(R_2|B_1)P(B_1)}{P(R_2|B_1)P(B_1)+P(R_2|R_1)P(R_1)}\\&=\frac{[r/(b+r+c)][b/(b+r)]}{[r/(b+r+c)][b/(b+r)]+[(r+c)/(b+r+c)][r/(b+r)]}\\&=\frac{b}{b+r+c}\end{aligned}$$

6. 令 B 表示两张牌都是 A 的事件.

(a) $P\{B|\text{黑桃 A}\}=\frac{P\{B,\text{黑桃 A}\}}{P\{\text{黑桃 A}\}}=\frac{\binom{1}{1}\binom{3}{1}}{\binom{52}{2}}\Bigg/\frac{\binom{1}{1}\binom{51}{1}}{\binom{52}{2}}=3/51$

(b) 因为第二张牌等可能地是剩下的 51 张牌中的任一张，其中三张是 A，所以可以得到这种情况下的概率

也是 3/51.

(c) 因为总是首先考虑互换哪张牌，其次考虑另一张牌，正如(b)中一样，结果应该是相同的．更一般的描述如下：

$$P\{B|\text{第二张是 A}\}=\frac{P\{B,\text{第二张是 A}\}}{P\{\text{第二张是 A}\}}=\frac{P(B)}{P(B)+P\{\text{第一张不是 A},\text{第二张是 A}\}}$$

$$=\frac{(4/52)(3/51)}{(4/52)(3/51)+(48/52)(4/51)}=3/51$$

(d) $P\{B|\text{至少一张是 A}\}=\dfrac{P(B)}{P\{\text{至少一张是 A}\}}=\dfrac{(4/52)(3/51)}{1-(48/52)(47/51)}=1/33$

7. $\dfrac{P(H|E)}{P(G|E)}=\dfrac{P(HE)}{P(GE)}=\dfrac{P(H)P(E|H)}{P(G)P(E|G)}$

假设 H 更可能发生，它是假设 G 发生的可能性的 1.5 倍.

8. 令 A 表示花活着的事件，令 W 表示花被浇的事件.

(a) $P(A)=P(A|W)P(W)+P(A|W^c)P(W^c)=(0.85)(0.9)+(0.2)(0.1)=0.785$

(b) $P(W^c|A^c)=\dfrac{P(A^c|W^c)P(W^c)}{P(A^c)}=\dfrac{(0.8)(0.1)}{0.215}=\dfrac{16}{43}$

9. 因为黑色老鼠有一个棕色的同胞，故它们的父母都有 1 个黑基因和 1 个棕基因．

(a) $P\{2\text{ 个黑基因}|\text{至少有 1 个黑基因}\}=\dfrac{P(2)}{P(\text{至少有 1 个黑基因})}=\dfrac{1/4}{3/4}=\dfrac{1}{3}$

(b) 令 F 表示 5 个后代都是黑色老鼠的事件．令 B_2 表示黑色老鼠有 2 个黑基因的事件，令 B_1 表示黑色老鼠有 1 个黑基因和 1 个棕基因的事件．

$$P(B_2|F)=\frac{P(F|B_2)P(B_2)}{P(F|B_2)P(B_2)+P(F|B_1)P(B_1)}=\frac{(1)(1/3)}{(1)(1/3)+(1/2)^5(2/3)}=\frac{16}{17}$$

10. 令 F 表示当前电流从 A 流到 B 的事件，令 C_i 表示继电器 i 闭合的事件，则有

$$P(F)=P(F|C_1)p_1+P(F|C_1^c)(1-p_1)$$

因为

$$P(F|C_1)=P(C_4\cup C_2C_5)=P(C_4)+P(C_2C_5)-P(C_4C_2C_5)=p_4+p_2p_5-p_4p_2p_5$$

又因

$$P(F|C_1^c)=P(C_2C_5\cup C_2C_3C_4)=p_2p_5+p_2p_3p_4-p_2p_3p_4p_5$$

故对(a)得到

$$P(F)=p_1(p_4+p_2p_5-p_4p_2p_5)+(1-p_1)p_2(p_5+p_3p_4-p_3p_4p_5)$$

对(b)，令 $q_i=1-p_i$，则

$$P(C_3|F)=P(F|C_3)P(C_3)/P(F)=p_3[1-P(C_1^cC_2^c\cup C_4^cC_5^c)]/P(F)$$

$$=p_3(1-q_1q_2-q_4q_5+q_1q_2q_4q_5)/P(F)$$

11. 令 A 表示部件 1 工作的事件，令 F 表示系统正常工作的事件．

(a) $P(A|F)=\dfrac{P(AF)}{P(F)}=\dfrac{P(A)}{P(F)}=\dfrac{1/2}{1-(1/2)^2}=\dfrac{2}{3}$

此处 $P(F)$ 可以计算，只要注意到它等于 1 减去部件 1 和部件 2 都失效的概率．

(b) $P(A|F)=\dfrac{P(AF)}{P(F)}=\dfrac{P(F|A)P(A)}{P(F)}=\dfrac{(3/4)(1/2)}{(1/2)^3+3(1/2)^3}=\dfrac{3}{4}$

此处 $P(F)$ 可以计算，只要注意到它等于 3 个部件都工作的概率减去只有两个部件工作的概率．

12. 如果接受连续旋转的结果是独立的，那么前 10 次旋转都出现黑数字的条件下，下一个结果的条件概率是不变的．

13. 以首次抛出的结果为条件：

$$P(\text{A 为奇数})=P_1(1-P_2)(1-P_3)+(1-P_1)P_2P_3+P_1P_2P_3P(\text{A 为奇数})$$

$$+(1-P_1)(1-P_2)(1-P_3)P(\text{A 为奇数})$$

故

$$P(\text{A 为奇数})=\frac{P_1(1-P_2)(1-P_3)+(1-P_1)P_2P_3}{P_1+P_2+P_3-P_1P_2-P_1P_3-P_2P_3}$$

14. 令 A 和 B 分别表示首次试验结果较大的事件和第二次试验结果较大的事件，并且令 E 表示试验结果相等的事件，则

$$1=P(A)+P(B)+P(E)$$

但是由对称性，$P(A)=P(B)$，因此，

$$P(B)=\frac{1-P(E)}{2}=\frac{1-\sum_{i=1}^{n}p_i^2}{2}$$

解此问题的另一种方法是，注意到

$$P(B)=\sum_i\sum_{j>i}P(\text{首次试验结果为 } i,\text{第二次试验结果为 } j)=\sum_i\sum_{j>i}p_ip_j$$

为了说明上面得到的 $P(B)$ 是相等的，注意到

$$1=\sum_{i=1}^{n}p_i\sum_{j=1}^{n}p_j=\sum_i\sum_j p_ip_j=\sum_i p_i^2+\sum_i\sum_{j\neq i}p_ip_j=\sum_i p_i^2+2\sum_i\sum_{j>i}p_ip_j$$

15. 令 $E=\{\text{A 比 B 得到更多的正面}\}$，则

$$\begin{aligned}P(E)=&P(E|\text{双方抛 } n \text{ 次后 A 得到更多的正面})P(\text{双方抛 } n \text{ 次后 A 得到更多的正面})\\&+P(E|\text{双方抛 } n \text{ 次后平})P(\text{双方抛 } n \text{ 次后平})\\&+P(E|\text{双方抛 } n \text{ 次后 B 得到更多的正面})P(\text{双方抛 } n \text{ 次后 B 得到更多的正面})\\=&P(\text{A 得到更多的正面})+\frac{1}{2}P(\text{平})\end{aligned}$$

由对称性，

$$P(\text{A 得到更多的正面})=P(\text{B 得到更多的正面})=\frac{1-P(\text{平})}{2}$$

因此，

$$P(E)=\frac{1}{2}$$

16. (a) 不真：掷 2 个骰子，令 $E=\{\text{点数之和是 }7\}$，$F=\{\text{第一个骰子点数不是 }4\}$，$G=\{\text{第二个骰子点数不是 }3\}$，那么

$$P(E|F\cup G)=\frac{P\{7,\text{不是}(4,3)\}}{P\{\text{不是}(4,3)\}}=\frac{5/36}{35/36}=5/35\neq P(E)$$

(b)
$$\begin{aligned}P(E(F\cup G))&=P(EF\cup EG)\\&=P(EF)+P(EG)\qquad \text{因为 } EFG=\varnothing\\&=P(E)[P(F)+P(G)]\\&=P(E)P(F\cup G)\qquad \text{因为 } GF=\varnothing\end{aligned}$$

(c)
$$\begin{aligned}P(G|EF)&=\frac{P(EFG)}{P(EF)}\\&=\frac{P(E)P(FG)}{P(EF)}\qquad \text{因为 } E \text{ 和 } FG \text{ 是独立的}\\&=\frac{P(E)P(F)P(G)}{P(E)P(F)}\qquad \text{由独立性}\\&=P(G)\end{aligned}$$

17. (a) 必然假；如果它们是互不相容的，则

$$0=P(AB)\neq P(A)P(B)$$

(b) 必然假；如果它们是独立的，则

$$P(AB)=P(A)P(B)>0$$

(c) 必然假；如果它们互不相容，则

$$P(A\cup B)=P(A)+P(B)=1.2$$

(d) 可能真.

18. (a)，(b)，(c)中的概率分别为:0.5，$(0.8)^3=0.512$，$(0.9)^7\approx 0.4783$.

19. 令 $D_i(i=1,2)$表示收音机 i 有问题的事件．令 A 和 B 分别表示收音机由工厂 A 和工厂 B 生产的事件，那么

$$P(D_2|D_1)=\frac{P(D_1D_2)}{P(D_1)}=\frac{P(D_1D_2|A)P(A)+P(D_1D_2|B)P(B)}{P(D_1|A)P(A)+P(D_1|B)P(B)}$$

$$=\frac{(0.05)^2(1/2)+(0.01)^2(1/2)}{(0.05)(1/2)+(0.01)(1/2)}=13/300$$

20. 给定 $P(AB)=P(B)$，一定可以证明 $P(B^cA^c)=P(A^c)$.

一种方法如下：

$$P(B^cA^c)=P((A\cup B)^c)=1-P(A\cup B)=1-P(A)-P(B)+P(AB)=1-P(A)=P(A^c)$$

21. 对 $n=0$ 结论是成立的．用 A_i 表示经过 n 次取球后，箱中有 i 个红球的事件，假设

$$P(A_i)=\frac{1}{n+1},\quad i=1,\cdots,n+1$$

令 B_j 表示经过 $n+1$ 次取球后，箱中有 j 个红球的事件，其中 $j=1,\cdots,n+2$，那么

$$P(B_j)=\sum_{i=1}^{n+1}P(B_j\mid A_i)P(A_i)=\frac{1}{n+1}\sum_{i=1}^{n+1}P(B_j\mid A_i)=\frac{1}{n+1}[P(B_j\mid A_{j-1})+P(B_j\mid A_j)]$$

因为经过 n 次取球后箱中有 $n+2$ 个球，注意到 $P(B_j|A_{j-1})$表示箱中的 $n+2$ 个球中有 $j-1$ 个红球时，取得红球的概率；$P(B_j|A_j)$表示箱中的 $n+2$ 球中有 j 个红球时，取得红球的概率．因此，

$$P(B_j|A_{j-1})=\frac{j-1}{n+2},\quad P(B_j|A_j)=\frac{n+2-j}{n+2}$$

将上面的等式代入 $P(B_j)$得到

$$P(B_j)=\frac{1}{n+1}\left[\frac{j-1}{n+2}+\frac{n+2-j}{n+2}\right]=\frac{1}{n+2}$$

用归纳法证毕．

22. 如果 A_i 表示第 i 人得到一张 A 的事件，则

$$P(A_i)=1-\frac{\binom{2n-2}{n}}{\binom{2n}{n}}=1-\frac{1}{2}\frac{n-1}{2n-1}=\frac{3n-1}{4n-2}$$

通过给这些 A 编号，且没有得到第一个 A 的人将得到剩下的 $2n-1$ 张牌中的 n 张，由此得到

$$P(A_1A_2)=\frac{n}{2n-1}$$

因此，

$$P(A_2^c|A_1)=1-P(A_2|A_1)=1-\frac{P(A_1A_2)}{P(A_1)}=\frac{n-1}{3n-1}$$

把牌的划分结果作为两个试验的结果，此处如果第一个人得到编号为 i 的牌 A，则认为试验 $i(i=1,2)$ 成功．因为两个 A 的位置是独立的，当 n 趋于无穷时，每一个 A 等可能地给任一个人，试验变成是独立的，每一个试验成功的概率为 1/2. 因此，在 $n\to\infty$的极限情况下，这个问题变成一个确定两个正面出现的条件概率问题，条件是:当抛两枚均匀的硬币时，至少有一个正面出现．因为$\frac{n-1}{3n-1}$收敛到 1/3，所以答案符合例 2d 的结论．

第 4 章

1. 因为概率和为 1，必有 $4P\{X=3\}+0.5=1$，这隐含着 $P\{X=0\}=0.375$，$P\{X=3\}=0.125$. 因此，

$E[X]=1(0.3)+2(0.2)+3(0.125)=1.075.$

2. 这个关系暗含着 $p_i=c^i p_0, i=1,2$，此处 $p_i=P\{X=i\}$. 因为这些概率的和为 1，可以得到

$$p_0(1+c+c^2)=1 \Rightarrow p_0=\frac{1}{1+c+c^2}$$

因此

$$E[X]=p_1+2p_2=\frac{c+2c^2}{1+c+c^2}$$

3. 令 X 表示抛的次数，那么 X 的概率质量函数为

$$p_2=p^2+(1-p)^2, p_3=1-p_2=2p(1-p)$$

因此，

$$E[X]=2p_2+3p_3=2p_2+3(1-p_2)=3-p^2-(1-p)^2$$

4. 随机选择一个有 i 个孩子的家庭的概率为 n_i/m. 因此，

$$E[X]=\sum_{i=1}^{r} in_i/m$$

并且，由于有 i 个孩子的家庭共有 in_i 个孩子，因此从一个有 i 个孩子的家庭中随机选择一个孩子的概率为 $in_i/\sum_{i=1}^{r} in_i$. 于是，

$$E[Y]=\frac{\sum_{i=1}^{r} i^2 n_i}{\sum_{i=1}^{r} in_i}$$

因此，必有

$$\frac{\sum_{i=1}^{r} i^2 n_i}{\sum_{i=1}^{r} in_i} \geqslant \frac{\sum_{i=1}^{r} in_i}{\sum_{i=1}^{r} n_i}$$

或者等价地，

$$\sum_{j=1}^{r} n_j \sum_{i=1}^{r} i^2 n_i \geqslant \sum_{i=1}^{r} in_i \sum_{j=1}^{r} jn_j$$

也等价于

$$\sum_{i=1}^{r}\sum_{j=1}^{r} i^2 n_i n_j \geqslant \sum_{i=1}^{r}\sum_{j=1}^{r} ij n_i n_j$$

但是，对固定的一对 i,j，上式左端和中 $n_i n_j$ 的系数为 i^2+j^2，然而在右端和中它的系数为 $2ij$. 因此，充分说明

$$i^2+j^2 \geqslant 2ij$$

上式成立是因为$(i-j)^2 \geqslant 0$.

5. 令 $p=P\{X=1\}$. 那么 $E[X]=p$, $\mathrm{Var}(X)=p(1-p)$，同时

$$p=3p(1-p)$$

这暗含着 $p=2/3$. 因此，$P\{X=0\}=1/3$.

6. 如果你赌 x，这个数赌赢的概率为 p，赌输的概率为 $1-p$，那么你赢的期望值为

$$xp-x(1-p)=(2p-1)x$$

当且仅当 $p>1/2$ 时，上式是正的(且关于 x 递增). 因此，当 $p\leqslant 1/2$ 时，一个人最大的期望收益是赌 0. 当 $p>1/2$ 时，一个人最大的期望收益是最大的可能值．因此，如果信息是出现正面的概率为 0.6 的硬币被选择，则应该赌 10，如果信息是出现正面的概率为 0.3 的硬币被选择，则应该赌 0. 因此，你的期望收益为

$$\frac{1}{2}(1.2-1)10+\frac{1}{2}0-C=1-C$$

在没有信息的情况下，你的期望收益是0(因为在这种情况下赢的概率为$\frac{1}{2}(0.6)+\frac{1}{2}(0.3)<1/2$)，如果信息价值小于1，那么购买它有利.

7. (a) 如果你翻阅红纸，并且观察到值 x，那么如果你换为蓝纸，期望收益是

$$2x(1/2)+x/2(1/2)=5x/4>x$$

因此，转换总是较好的.

(b) 假设慈善家在红纸上写下值 x，故蓝纸上的数为 $2x$ 或者 $x/2$. 注意如果 $x/2\geqslant y$，那么蓝纸上的数将至少是 y. 因此被接受. 因此，在这种情况下，奖金等可能地是 $2x$ 或者 $x/2$，因此

$$E[R_y(x)]=5x/4,\text{ 如果 } x/2\geqslant y$$

如果 $x/2<y\leqslant 2x$，那么如果值是 $2x$，蓝纸将被接受，如果值是 $x/2$，蓝纸被拒绝. 因此，

$$E[R_y(x)]=2x(1/2)+x(1/2)=3x/2,\text{ 如果 } x/2<y\leqslant 2x$$

最后，如果 $2x<y$，那么蓝纸将被拒绝. 因此，这种情况下，报酬是 x，于是，

$$R_y(x)=x,\text{ 如果 } 2x<y$$

也就是说，当数 x 写在红纸上时，在 y—策略下，证明了期望收益是

$$E[R_y(x)]=\begin{cases}x & \text{若 } x<y/2\\ 3x/2 & \text{若 } y/2\leqslant x<2y\\ 5x/4 & \text{若 } x\geqslant 2y\end{cases}$$

8. 假设做 n 次独立试验，每次试验成功的概率为 p. 那么当且仅当失败次数大于等于 $n-i$ 时，成功次数小于等于 i. 但是由于每次试验失败的概率都为 $1-p$，故失败次数是参数为 n 和$1-p$的二项随机变量. 因此，

$$P\{B(n,p)\leqslant i\}=P\{B(n,1-p)\geqslant n-i\}=1-P\{B(n,1-p)\leqslant n-i-1\}$$

最后的等式成立是因为失败次数大于等于 $n-i$ 的概率等于 1 减去失败次数小于 $n-i$ 的概率.

9. 由于 $E[X]=np$, $\mathrm{Var}(X)=np(1-p)$，我们给出 $np=6$, $np(1-p)=2.4$. 从而，$1-p=0.4$ 或 $p=0.6$，$n=10$. 因此，

$$P\{X=5\}=\binom{10}{5}(0.6)^5(0.4)^5$$

10. 令 X_i 表示取出第 i 个球，$i=1,\cdots,m$，则

$$P\{X\leqslant k\}=P\{X_1\leqslant k,X_2\leqslant k,\cdots,X_m\leqslant k\}=P\{X_1\leqslant k\}P\{X_2\leqslant k\}\cdots P\{X_m\leqslant k\}=\left(\frac{k}{n}\right)^m$$

因此，

$$P\{X=k\}=P\{X\leqslant k\}-P\{X\leqslant k-1\}=\left(\frac{k}{n}\right)^m-\left(\frac{k-1}{n}\right)^m$$

11. (a) 给定甲赢得第 1 场比赛，如果在乙赢得 3 场比赛前，甲连续再赢 2 场比赛，则甲赢. 因而

$$P\{\text{甲赢}\mid\text{甲赢得第 1 场比赛}\}=\sum_{i=2}^{4}\binom{4}{i}p^i(1-p)^{4-i}$$

(b)
$$P\{\text{甲赢得第 1 场比赛}\mid\text{甲赢}\}=\frac{P\{\text{甲赢}\mid\text{甲赢得第 1 场比赛}\}P\{\text{甲赢得第 1 场比赛}\}}{P\{\text{甲赢}\}}=\frac{\sum_{i=2}^{4}\binom{4}{i}p^{i+1}(1-p)^{4-i}}{\sum_{i=3}^{5}\binom{5}{i}p^i(1-p)^{5-i}}$$

12. 以这个周末该队是否赢为条件，获得解答：

$$0.5\sum_{i=3}^{4}\binom{4}{i}(0.4)^i(0.6)^{4-i}+0.5\sum_{i=3}^{4}\binom{4}{i}(0.7)^i(0.3)^{4-i}$$

13. 假设狂风数可以近似为泊松随机变量，我们得到答案：

$$\sum_{i=0}^{3} e^{-5.2}(5.2)^i/i!$$

14. $E[Y] = \sum_{i=1}^{\infty} iP\{X = i\}/P\{X > 0\} = E[X]/P\{X > 0\} = \frac{\lambda}{1-e^{-\lambda}}$

15. (a) $\binom{8}{3}(9/19)^3(10/19)^5(9/19)=\binom{8}{3}(9/19)^4(10/19)^5$

(b) 如果 W 是他最后的获胜，X 是他下的赌注数，由于他赢了 4 次赌博，输了 $X-4$ 次赌博，那么

$$W=20-5(X-4)=40-5X$$

因此，

$$E[W]=40-5E[X]=40-5[4/(9/19)]=-20/9$$

16. 一次抛硬币之后没有出现“奇数人”的概率为 1/4，即所有的三枚硬币都出现同一面的概率．

(a) $(1/4)^2(3/4)=3/64$　　(b) $(1/4)^4=1/256$

17. 令 $q=1-p$.

$$E[1/X] = \sum_{i=1}^{\infty} \frac{1}{i}q^{i-1}p = \frac{p}{q}\sum_{i=1}^{\infty}q^i/i = \frac{p}{q}\sum_{i=1}^{\infty}\int_0^q x^{i-1}\mathrm{d}x$$

$$= \frac{p}{q}\int_0^q \sum_{i=1}^{\infty} x^{i-1}\mathrm{d}x = \frac{p}{q}\int_0^q \frac{1}{1-x}\mathrm{d}x = \frac{p}{q}\int_p^1 \frac{1}{y}\mathrm{d}y = -\frac{p}{q}\log(p)$$

18. 因为$\frac{X-b}{a-b}$以概率 p 等于 1，或者以概率 $1-p$ 等于 0，因此它是一个参数为 p 的伯努利随机变量．因为这样一个伯努利随机变量的方差为 $p(1-p)$，可以得到

$$p(1-p)=\mathrm{Var}\left(\frac{X-b}{a-b}\right)=\frac{1}{(a-b)^2}\mathrm{Var}(X-b)=\frac{1}{(a-b)^2}\mathrm{Var}(X)$$

因此，

$$\mathrm{Var}(X)=(a-b)^2p(1-p)$$

19. 令 X 表示比赛总局数，Y 表示输的局数．

(a) 第 4 次比赛后继续玩，直到你输．于是，$X-4$ 是一个参数为 $1-p$ 的几何随机变量，因此

$$E[X]=E[4+(X-4)]=4+E[X-4]=4+\frac{1}{1-p}$$

(b) 如果令 Z 表示在前 4 局比赛中输的局数，则 Z 是一个参数为 4 和 $1-p$ 的二项随机变量．因为$Y=Z+1$，我们有

$$E[Y]=E[Z+1]=E[Z]+1=4(1-p)+1$$

20. 在 m 个黑球之前将有 n 个白球取出，当且仅当在前 $n+m-1$ 次取球中至少有 n 个白球．(和第 3 章中例 4i 的分赌注问题比较．) 用 X 表示前 $n+m-1$ 次取出的球中的白球数，那么 X 是一个超几何随机变量，并且

$$P\{X \geqslant n\} = \sum_{i=n}^{n+m-1} P\{X = i\} = \sum_{i=n}^{n+m-1} \frac{\binom{N}{i}\binom{M}{n+m-1-i}}{\binom{N+M}{n+m-1}}$$

第 5 章

1. 令 X 表示上场时间(分钟).

(a) $P\{X>15\}=1-P\{X\leqslant 15\}=1-5(0.025)=0.875$

(b) $P\{20<X<35\}=10(0.05)+5(0.025)=0.625$

(c) $P\{X<30\}=10(0.025)+10(0.05)=0.75$

(d) $P\{X>36\}=4(0.025)=0.1$

2. (a) $1=\int_0^1 cx^n \mathrm{d}x = c/(n+1) \Rightarrow c = n+1$

(b) $P\{X>x\} = (n+1)\int_x^1 x^n \mathrm{d}x = x^{n+1}\Big|_x^1 = 1-x^{n+1}$

3. 首先通过下式得到 c 的值：

$$1=\int_0^2 cx^4 \mathrm{d}x = 32c/5 \Rightarrow c = 5/32$$

(a) $E[X]=\frac{5}{32}\int_0^2 x^5 \mathrm{d}x = \frac{5}{32}\frac{64}{6} = 5/3$

(b) $E[X^2]=\frac{5}{32}\int_0^2 x^6 \mathrm{d}x = \frac{5}{32}\frac{128}{7} = 20/7 \Rightarrow \mathrm{Var}(X) = 20/7-(5/3)^2 = 5/63$

4. 由

$$1=\int_0^1 (ax+bx^2)\mathrm{d}x = a/2+b/3$$

$$0.6=\int_0^1 (ax^2+bx^3)\mathrm{d}x = a/3+b/4$$

得到 $a=3.6, b=-2.4$. 因此，

(a) $P\{X<1/2\}=\int_0^{1/2}(3.6x-2.4x^2)\mathrm{d}x = (1.8x^2-0.8x^3)\Big|_0^{1/2} = 0.35$

(b) $E[X^2]=\int_0^1(3.6x^3-2.4x^4)\mathrm{d}x = 0.42 \Rightarrow \mathrm{Var}(X) = 0.06$

5. 对 $i=1,\cdots,n$

$$P\{X=i\}=P\{\mathrm{Int}(nU)=i-1\}=P\{i-1\leqslant nU<i\}=P\left\{\frac{i-1}{n}\leqslant U<\frac{i}{n}\right\}=1/n$$

6. 如果你出价 x，$70\leqslant x\leqslant 140$，那么你将赢得竞标，并以概率 $(140-x)/70$ 获利 $x-100$，否则，你将失去竞标，获得 0 利润．因此，如果你出价 x，你的期望利润是

$$\frac{1}{70}(x-100)(140-x)=\frac{1}{70}(240x-x^2-14\,000)$$

微分并令上式等于 0，得到

$$240-2x=0$$

因此，你应该出价 12 万美元，期望利润将是 4/7 万美元．

7. (a) $P\{U>0.1\}=9/10$

(b) $P\{U>0.2|U>0.1\}=P\{U>0.2\}/P\{U>0.1\}=8/9$

(c) $P\{U>0.3|U>0.2, U>0.1\}=P\{U>0.3\}/P\{U>0.2\}=7/8$

(d) $P\{U>0.3\}=7/10$

(e) 部分可以通过(a)，(b)，(c)中的概率相乘得到．

8. 令 X 表示测验值，令 $Z=(X-100)/15$. 注意 Z 是一个标准正态随机变量．

(a) $P\{X>125\}=P\{Z>25/15\}\approx 0.047\,8$

(b) $P\{90<X<110\}=P\{-10/15<Z<10/15\}=P\{Z<2/3\}-P\{Z<-2/3\}$
$=P\{Z<2/3\}-[1-P\{Z<2/3\}]\approx 0.495\,0$

9. 令 X 表示所花费的时间．想要得到满足下式的 x：

$$P\{X>x\}=0.05$$

上式等价于

$$P\left\{\frac{X-40}{7}>\frac{x-40}{7}\right\}=0.05$$

也就是说，需要找到满足下式的 x：

$$P\left\{Z>\frac{x-40}{7}\right\}=0.05$$

此处 Z 是一个标准正态随机变量．但是

$$P\{Z>1.645\}=0.05$$

于是

$$\frac{x-40}{7}=1.645 \quad 或者 \quad x=51.515$$

因此你最晚应在 12:00 后 8.485 分钟从家中出发．

10. 令 X 表示轮胎的寿命(以一千英里为单位)，令 $Z=(X-34)/4$. 注意 Z 是一个标准正态随机变量．

(a) $P\{X>40\}=P\{Z>1.5\}\approx 0.0668$

(b) $P\{30<X<35\}=P\{-1<Z<0.25\}=P\{Z<0.25\}-P\{Z>1\}\approx 0.44$

(c) $P\{X>40|X>30\}=P\{X>40\}/P\{X>30\}=P\{Z>1.5\}/P\{Z>-1\}\approx 0.079$

11. 令 X 表示明年的降雨量，令 $Z=(X-40.2)/8.4$.

(a) $P\{X>44\}=P\{Z>3.8/8.4\}\approx P\{Z>0.4524\}\approx 0.3255$

(b) $\binom{7}{3}(0.3255)^3(0.6745)^4$

12. 以一千美元为单位，令 M_i 和 W_i 分别表示每年收入至少为 i 个单位的男士和女士数，并且令 Z 是一个标准正态随机变量．

(a) $P\{W_{25}\geqslant 70\}=P\{W_{25}\geqslant 69.5\}$

$$=P\left\{\frac{W_{25}-200(0.34)}{\sqrt{200(0.34)(0.66)}}\geqslant\frac{69.5-200(0.34)}{\sqrt{200(0.34)(0.66)}}\right\}$$

$$\approx P\{Z\geqslant 0.2239\}\approx 0.4114$$

(b) $P\{M_{25}\leqslant 120\}=P\{M_{25}\leqslant 120.5\}$

$$=P\left\{\frac{M_{25}-200(0.587)}{\sqrt{200(0.587)(0.413)}}\leqslant\frac{120.5-200(0.587)}{\sqrt{200(0.587)(0.413)}}\right\}$$

$$\approx P\{Z\leqslant 0.4452\}\approx 0.6719$$

(c) $P\{M_{20}\geqslant 150\}=P\{M_{20}\geqslant 149.5\}$

$$P\left\{\frac{M_{20}-200(0.745)}{\sqrt{200(0.745)(0.255)}}\geqslant\frac{149.5-200(0.745)}{\sqrt{200(0.745)(0.255)}}\right\}$$

$$\approx P\{Z\geqslant 0.0811\}\approx 0.4677$$

$P\{W_{20}\geqslant 100\}=P\{W_{20}\geqslant 99.5\}$

$$=P\left\{\frac{W_{20}-200(0.534)}{\sqrt{200(0.534)(0.466)}}\geqslant\frac{99.5-200(0.534)}{\sqrt{200(0.534)(0.466)}}\right\}$$

$$\approx P\{Z\geqslant -1.0348\}\approx 0.8496$$

因此，

$$P\{M_{20}\geqslant 150\}P\{W_{20}\geqslant 100\}\approx 0.3974$$

13. 由指数分布的无记忆性，可得结果 $e^{-4/5}$.

14. (a) $e^{-2^2}=e^{-4}$　　(b) $F(3)-F(1)=e^{-1}-e^{-9}$

(c) $\lambda(t)=2te^{-t^2}/e^{-t^2}=2t$

(d) 令 Z 是一标准正态随机变量．运用等式 $E[X]=\int_0^\infty P\{X>x\}dx$ 得到

$$E[X]=\int_0^\infty e^{-x^2}dx=2^{-1/2}\int_0^\infty e^{-y^2/2}dy=2^{-1/2}\sqrt{2\pi}P\{Z>0\}=\sqrt{\pi}/2$$

(e) 运用理论练习 5 的结论得到

$$E[X^2]=\int_0^{\infty}2xe^{-x^2}\,dx=-e^{-x^2}\Big|_0^{\infty}=1$$

故得 $\text{Var}(X)=1-\pi/4$.

15. (a) $P\{X>6\}=\exp\left\{-\int_0^6\lambda(t)dt\right\}=e^{-3.45}$

(b) $P\{X<8|X>6\}=1-P\{X>8|X>6\}=1-P\{X>8\}/P\{X>6\}$

$=1-e^{-5.65}/e^{-3.45}\approx 0.8892$

16. 对 $x\geqslant 0$

$$F_{1/X}(x)=P\{1/X\leqslant x\}=P\{X\leqslant 0\}+P\{X\geqslant 1/x\}=1/2+1-F_X(1/x)$$

微分得

$$f_{1/X}(x)=x^{-2}f_X(1/x)=\frac{1}{x^2\pi(1+(1/x)^2)}=f_X(x)$$

当 $x<0$ 时，证明过程类似.

17. 如果 X 表示前 n 次赌博中你获胜的次数，那么 n 次赌博后你的获胜数为

$$35X-(n-X)=36X-n$$

因而，想确定

$$p=P\{36X-n>0\}=P\{X>n/36\}$$

此处 X 是服从参数为 n 和 $p=1/38$ 的二项分布的随机变量.

(a) 当 $n=34$ 时，

$$\begin{aligned}p&=P\{X\geqslant 1\}\\&=P\{X>0.5\}\qquad\text{(连续性修正)}\\&=P\left\{\frac{X-34/38}{\sqrt{34(1/38)(37/38)}}>\frac{0.5-34/38}{\sqrt{34(1/38)(37/38)}}\right\}\\&=P\left\{\frac{X-34/38}{\sqrt{34(1/38)(37/38)}}>-0.4229\right\}\\&\approx\Phi(0.4229)\\&\approx 0.6638\end{aligned}$$

(如果你至少赢了一次赌博，那么 34 次赌博后你将领先，这种情况下的确切概率为 $1-(37/38)^{34}=0.5961$.)

(b) 当 $n=1\,000$ 时，

$$\begin{aligned}p&=P\{X>27.5\}\\&=P\left\{\frac{X-1\,000/38}{\sqrt{1\,000(1/38)(37/38)}}>\frac{27.5-1\,000/38}{\sqrt{1\,000(1/38)(37/38)}}\right\}\\&\approx 1-\Phi(0.2339)\\&\approx 0.4075\end{aligned}$$

这个确切的概率是 0.396 1，也即是服从参数为 $n=1\,000$，$p=1/38$ 的二项分布的随机变量大于 27 的概率.

(c) 当 $n=100\,000$ 时，

$$\begin{aligned}p&=P\{X>2\,777.5\}\\&=P\left\{\frac{X-100\,000/38}{\sqrt{100\,000(1/38)(37/38)}}>\frac{2\,777.5-100\,000/38}{\sqrt{100\,000(1/38)(37/38)}}\right\}\\&\approx 1-\Phi(2.883)\\&\approx 0.0020\end{aligned}$$

在这种情况下的确切概率为 0.002 1.

18. 如果 X 表示电池的寿命，那么所求的概率 $P\{X>s+t|X>t\}$ 由下式确定：

$$\begin{aligned}P\{X>s+t|X>t\}&=\frac{P\{X>s+t,X>t\}}{P\{X>t\}}\\&=\frac{P\{X>s+t\}}{P\{X>t\}}\\&=\frac{P\{X>s+t|\text{电池是型号 }1\}p_1+P\{X>s+t|\text{电池是型号 }2\}p_2}{P\{X>t|\text{电池是型号 }1\}p_1+P\{X>t|\text{电池是型号 }2\}p_2}\\&=\frac{\mathrm{e}^{-\lambda_1(s+t)}p_1+\mathrm{e}^{-\lambda_2(s+t)}p_2}{\mathrm{e}^{-\lambda_1t}p_1+\mathrm{e}^{-\lambda_2t}p_2}\end{aligned}$$

另一种方法是直接以电池的型号为条件，然后利用指数分布的无记忆性．也就是说，我们可以如下完成：

$$\begin{aligned}P\{X>s+t|X>t\}&=P\{X>s+t|X>t,\text{型号 }1\}P\{\text{型号 }1|X>t\}+P\{X>s+t|X>t,\text{型号 }2\}P\{\text{型号 }2|X>t\}\\&=\mathrm{e}^{-\lambda_1s}P\{\text{型号 }1|X>t\}+\mathrm{e}^{-\lambda_2s}P\{\text{型号 }2|X>t\}\end{aligned}$$

对 $i=1,2$，运用

$$\begin{aligned}P\{\text{型号 }i|X>t\}&=\frac{P\{\text{型号 }i,X>t\}}{P\{X>t\}}\\&=\frac{P\{X>t|\text{型号 }i\}p_i}{P\{X>t|\text{型号 }1\}p_1+P\{X>t|\text{型号 }2\}p_2}\\&=\frac{\mathrm{e}^{-\lambda_it}p_i}{\mathrm{e}^{-\lambda_1t}p_1+\mathrm{e}^{-\lambda_2t}p_2}\end{aligned}$$

第 6 章

1. (a) $3C+6C=1\Rightarrow C=1/9$

(b) 令 $p(i,j)=P\{X=i,Y=j\}$，那么

$$p(1,1)=4/9,p(1,0)=2/9,p(0,1)=1/9,p(0,0)=2/9$$

(c) $\frac{(12)!}{2^6}(1/9)^6(2/9)^6$

(d) $\frac{(12)!}{(4!)^3}(1/3)^{12}$

(e) $\sum_{i=8}^{12}\binom{12}{i}(2/3)^i(1/3)^{12-i}$

2. (a) 由 $p_j=P\{XYZ=j\}$，有

$$p_6=p_2=p_4=p_{12}=1/4$$

因此，

$$E[XYZ]=(6+2+4+12)/4=6$$

(b) 由 $q_j=P\{XY+XZ+YZ=j\}$，有

$$q_{11}=q_5=q_8=q_{16}=1/4$$

因此，

$$E[XY+XZ+YZ]=(11+5+8+16)/4=10$$

3. 下面将用到等式

$$\int_0^\infty \mathrm{e}^{-x}x^n\mathrm{d}x=n!$$

上式成立是因为 $\mathrm{e}^{-x}x^n/n!$ $(x>0)$是参数为 $n+1$ 和 λ 的 Γ 分布的密度函数，故有积分为 1.

(a) $1=C\int_0^\infty \mathrm{e}^{-y}\int_{-y}^{y}(y-x)\mathrm{d}x\mathrm{d}y=C\int_0^\infty \mathrm{e}^{-y}2y^2\mathrm{d}y=4C$

因此，$C=1/4$.

(b) 仅当 $y>x$ 和 $y>-x$ 时，联合密度函数不为 0，对 $x>0$，有

$$f_X(x)=\frac{1}{4}\int_x^\infty(y-x)\mathrm{e}^{-y}\mathrm{d}y=\frac{1}{4}\int_0^\infty u\mathrm{e}^{-(x+u)}\mathrm{d}u=\frac{1}{4}\mathrm{e}^{-x}$$

对 $x<0$，

$$f_X(x)=\frac{1}{4}\int_{-x}^{\infty}(y-x)e^{-y}dy=\frac{1}{4}\left[-ye^{-y}-e^{-y}+xe^{-y}\right]_{-x}^{\infty}=(-2xe^{x}+e^{x})/4$$

(c) $f_Y(y)=\frac{1}{4}e^{-y}\int_{-y}^{y}(y-x)dx=\frac{1}{2}y^2e^{-y}$

(d) $E[X]=\frac{1}{4}\left[\int_0^{\infty}xe^{-x}dx+\int_{-\infty}^{0}(-2x^2e^{x}+xe^{x})dx\right]$

$$=\frac{1}{4}\left[1-\int_0^{\infty}(2y^2e^{-y}+ye^{-y})dy\right]=\frac{1}{4}[1-4-1]=-1$$

(e) $E[Y]=\frac{1}{2}\int_0^{\infty}y^3e^{-y}dy=3$

4. (a) 令 $p_j=P\{XYZ=j\}$，有

$$p_1=1/8,\quad p_2=3/8,\quad p_4=3/8,\quad p_8=1/8$$

(b) 令 $p_j=P\{XY+XZ+YZ=j\}$，有

$$p_3=1/8,\quad p_5=3/8,\quad p_8=3/8,\quad p_{12}=1/8$$

(c) 令 $p_j=P\{X^2+YZ=j\}$，有

$$p_2=1/8,\quad p_3=1/4,\quad p_5=1/4,\quad p_6=1/4,\quad p_8=1/8$$

5. (a) $1=\int_0^1\int_1^5(x/5+cy)dydx=\int_0^1(4x/5+12c)dx=12c+2/5$

因此，$c=1/20$.

(b) 不独立，密度函数不能写成两个因子的乘积.

(c) $P\{X+Y>3\}=\int_0^1\int_{3-x}^5(x/5+y/20)dydx$

$$=\int_0^1[(2+x)x/5+25/40-(3-x)^2/40]dx$$

$$=1/5+1/15+5/8-19/120=11/15$$

6. (a) 独立，联合密度函数可以写成两个因子的乘积.

(b) $f_X(x)=x\int_0^2 ydy=2x,0<x<1$

(c) $f_Y(y)=y\int_0^1 xdx=y/2,0<y<2$

(d) $P\{X<x,Y<y\}=P\{X<x\}P\{Y<y\}=\min(1,x^2)\min(1,y^2/4),x>0,y>0$

(e) $E[Y]=\int_0^2 y^2/2dy=4/3$

(f) $P\{X+Y<1\}=\int_0^1 x\int_0^{1-x}ydydx=\frac{1}{2}\int_0^1 x(1-x)^2dx=1/24$

7. 令 T_i 表示类型 i 冲击发生的时间，$i=1,2,3$. 对 $s>0,t>0$，

$$\begin{aligned}P\{X_1>s,X_2>t\}&=P\{T_1>s,T_2>t,T_3>\max(s,t)\}\\&=P\{T_1>s\}P\{T_2>t\}P\{T_3>\max(s,t)\}\\&=\exp\{-\lambda_1 s\}\exp\{-\lambda_2 t\}\exp\{-\lambda_3\max(s,t)\}\\&=\exp\{-(\lambda_1 s+\lambda_2 t+\lambda_3\max(s,t))\}\end{aligned}$$

8. (a) 不能，页面上有许多广告的比页面上有较少广告的更不容易被选择.

(b) $\frac{1}{m}\frac{n(i)}{n}$

(c) $\frac{\sum_{i=1}^{m}n(i)}{nm}=\bar{n}/n$，此处 $\bar{n}=\sum_{i=1}^{m}n(i)/m$

(d) $(1-\bar{n}/n)^{k-1}\frac{1}{m}\frac{n(i)}{n}\frac{1}{n(i)}=(1-\bar{n}/n)^{k-1}/(nm)$

(e) $\sum_{k=1}^{\infty}\frac{1}{nm}(1-\bar{n}/n)^{k-1}=\frac{1}{\bar{n}m}$

(f) 迭代数服从均值为 $n\sqrt{n}$ 的几何分布．

9. (a) $P\{X=i\}=1/m, i=1,\cdots,m.$

(b) 步骤 2. 产生一个随机变量 U，U 服从(0,1)上的均匀分布．如果 $U<n(X)/n$，转到步骤 3．否则转到步骤 1.

步骤 3. 产生一个随机变量 U，U 服从(0,1)上的均匀分布．在第 X 页上选择这个元素放在位置 $[n(X)U]+1$ 上．

10. 是的，它们是独立的．通过考虑 X_N 是否与 N 独立可以容易地得到．除了确实这样以外，还因为第一个随机变量大于 c 发生并不影响它取值的概率分布，即 $(c,1)$ 上的均匀分布．

11. 令 p_i 表示在一次投镖中获得 i 分，那么

$$p_{30}=\pi/36$$
$$p_{20}=4\pi/36-p_{30}=\pi/12$$
$$p_{10}=9\pi/36-p_{20}-p_{30}=5\pi/36$$
$$p_0=1-p_{10}-p_{20}-p_{30}=1-\pi/4$$

(a) $\pi/12$　　(b) $\pi/9$

(c) $1-\pi/4$　　(d) $\pi(30/36+20/12+50/36)=35\pi/9$

(e) $(\pi/4)^2$　　(f) $2(\pi/36)(1-\pi/4)+2(\pi/12)(5\pi/36)$

12. 令 Z 是一标准正态随机变量．

(a) $P\left\{\sum_{i=1}^{4}X_i>0\right\}=P\left\{\frac{\sum_{i=1}^{4}X_i-6}{\sqrt{24}}>\frac{-6}{\sqrt{24}}\right\}\approx P\{Z>-1.2247\}\approx 0.8897$

(b)
$$P\left\{\sum_{i=1}^{4}X_i>0\,\Big|\,\sum_{i=1}^{2}X_i=-5\right\}=P\{X_3+X_4>5\}$$
$$=P\left\{\frac{X_3+X_4-3}{\sqrt{12}}>2/\sqrt{12}\right\}$$
$$\approx P\{Z>0.5774\}\approx 0.2818$$

(c)
$$P\left\{\sum_{i=1}^{4}X_i>0\,\Big|\,X_1=5\right\}=P\{X_2+X_3+X_4>-5\}$$
$$=P\left\{\frac{X_2+X_3+X_4-4.5}{\sqrt{18}}>-9.5/\sqrt{18}\right\}$$
$$\approx P\{Z>-2.239\}\approx 0.9874$$

13. 在下列各项中，C 不依赖 n.

$$P\{N=n|X=x\}=f_{X|N}(x|n)P\{N=n\}/f_X(x)$$
$$=C\frac{1}{(n-1)!}(\lambda x)^{n-1}(1-p)^{n-1}$$
$$=C(\lambda(1-p)x)^{n-1}/(n-1)!$$

上式表明，在 $X=x$ 条件下，$N-1$ 是均值为 $\lambda(1-p)x$ 的泊松随机变量．也就是说，

$$P\{N=n|X=x\}=P\{N-1=n-1|X=x\}=e^{-\lambda(1-p)x}(\lambda(1-p)x)^{n-1}/(n-1)!, n\geq 1$$

14. (a) 变换的雅可比行列式为

$$J=\begin{vmatrix}1 & 0\\ 1 & 1\end{vmatrix}=1$$

由于等式 $u=x, v=x+y$ 暗含着 $x=u, y=v-u$，可以得到

$$f_{U,V}(u,v)=f_{X,Y}(u,v-u)=1, 0<u<1, 0<v-u<1$$

或等价地，

$$f_{U,V}(u,v)=1, \max(v-1,0)<u<\min(v,1)$$

(b) 对 $0<v<1$，

$$f_V(v)=\int_0^v du = v$$

对 $1\leqslant v\leqslant 2$，

$$f_V(v)=\int_{v-1}^1 du = 2-v$$

15. 令 U 是一个服从(7,11)上均匀分布的随机变量．如果你出价 x，$7\leqslant x\leqslant 10$，你将以下列概率是最高的出价者：

$$(P\{U<x\})^3=\left(P\left\{\frac{U-7}{4}<\frac{x-7}{4}\right\}\right)^3=\left(\frac{x-7}{4}\right)^3$$

因此，如果你出价 x，你的期望利润 $E[G(x)]$ 是

$$E[G(x)]=\frac{1}{4}(x-7)^3(10-x)$$

计算表明当 $x=37/4$ 时，期望利润最大．

16. 令 $i_1,i_2,\cdots,i_n$ 是 $1,2,\cdots,n$ 的一个排列．

$$P\{X_1=i_1,X_2=i_2,\cdots,X_n=i_n\}=P\{X_1=i_1\}P\{X_2=i_2\}\cdots P\{X_n=i_n\}=p_{i_1}p_{i_2}\cdots p_{i_n}=p_1p_2\cdots p_n$$

因此，所求的概率为 $n!\ p_1p_2\cdots p_n$，当所有的 $p_i=1/n$ 时，它变为 $\frac{n!}{n^n}$.

17. (a) 因为 $\sum_{i=1}^n X_i=\sum_{i=1}^n Y_i$，所以 $N=2M$.

(b) 考虑 $n-k$ 个坐标，它们的 Y 值都等于 0，称它们为红色坐标．因为 X 值等于 1 的 k 个坐标等可能地是含 k 个坐标的 $\binom{n}{k}$ 个集合中的任一个，所以在 k 个坐标中红色坐标数和下述情况有相同的分布：从 $n-k$ 个球是红球的 n 个球中随机选择 k 个球，选出的红球数．因此，M 是一个服从超几何分布的随机变量．

(c) $E[N]=E[2M]=2E[M]=\frac{2k(n-k)}{n}$

(d) 用第 4 章式(9-6)给出的超几何分布的方差公式，可得

$$\mathrm{Var}(N)=4\mathrm{Var}(M)=4\frac{n-k}{n-1}k(1-k/n)(k/n)$$

***18.** (a) 首先注意 $S_n-S_k=\sum_{i=k+1}^n Z_i$ 是服从均值为 0、方差为 $n-k$ 的正态随机变量，并且和 S_k 独立．因此，给定 $S_k=y$，S_n 是服从均值为 y、方差为 $n-k$ 的正态随机变量．

(b) 给定 $S_n=x$，S_k 的条件密度函数是变量为 y 的密度函数，任何不依赖于 y 的都认为是常数．（例如，x 被认为是一个固定的常数．）下面的量 $C_i(i=1,2,3,4)$ 都是不依赖于 y 的常数．

$$\begin{aligned}f_{S_k|S_n}(y|x)&=\frac{f_{S_k,S_n}(y,x)}{f_{S_n}(x)}\\&=C_1 f_{S_n|S_k}(x|y)f_{S_k}(y)\quad\left(\text{此处 } C_1=\frac{1}{f_{S_n}(x)}\right)\\&=C_1\frac{1}{\sqrt{2\pi}\sqrt{n-k}}e^{-(x-y)^2/2(n-k)}\frac{1}{\sqrt{2\pi}\sqrt{k}}e^{-y^2/2k}\end{aligned}$$

$$
\begin{aligned}
&= C_2 \exp\left\{-\frac{(x-y)^2}{2(n-k)} - \frac{y^2}{2k}\right\} \\
&= C_3 \exp\left\{\frac{2xy}{2(n-k)} - \frac{y^2}{2(n-k)} - \frac{y^2}{2k}\right\} \\
&= C_3 \exp\left\{-\frac{n}{2k(n-k)}\left(y^2 - 2\,\frac{k}{n}xy\right)\right\} \\
&= C_3 \exp\left\{-\frac{n}{2k(n-k)}\left[\left(y-\frac{k}{n}x\right)^2 - \left(\frac{k}{n}x\right)^2\right]\right\} \\
&= C_4 \exp\left\{-\frac{n}{2k(n-k)}\left(y-\frac{k}{n}x\right)^2\right\}
\end{aligned}
$$

我们知道上式是正态分布的密度函数，均值为$\frac{k}{n}x$，方差为$\frac{k(n-k)}{n}$.

第 7 章

1. (a) $d = \sum_{i=1}^{m} 1/n(i)$

(b) $P\{X=i\} = P\{[mU]=i-1\} = P\{i-1 \leqslant mU < i\} = 1/m, i=1,\cdots,m$

(c) $E\left[\frac{m}{n(X)}\right] = \sum_{i=1}^{m} \frac{m}{n(i)} P\{X=i\} = \sum_{i=1}^{m} \frac{m}{n(i)} \frac{1}{m} = d$

2. 如果第 j 次取出的是白球，且第 $j+1$ 次取出的是黑球，令 I_j 等于 1，否则令 I_j 等于 0. 如果 X 表示一个白球后是一个黑球这样的事件数，那么 X 可以表示为

$$X = \sum_{j=1}^{n+m-1} I_j$$

因而，

$$
\begin{aligned}
E[X] &= \sum_{j=1}^{n+m-1} E[I_j] \\
&= \sum_{j=1}^{n+m-1} P\{\text{第 } j \text{ 次选择白球,第 } j+1 \text{ 次选择黑球}\} \\
&= \sum_{j=1}^{n+m-1} P\{\text{第 } j \text{ 次选择白球}\} P\{\text{第 } j+1 \text{ 次选择黑球} \mid \text{第 } j \text{ 次选择白球}\} \\
&= \sum_{j=1}^{n+m-1} \frac{n}{n+m} \frac{m}{n+m-1} \\
&= \frac{nm}{n+m}
\end{aligned}
$$

上式用到了下面的事实：第 j 次选择的球等可能地是 $n+m$ 个球中的任一个，给定已选择的是一个白球的情况下，下一次选择的球等可能地是其他的 $n+m-1$ 个球中的任一个.

3. 任意给这些夫妇编号，如果编号为 j 的已婚夫妇坐在编号相同的桌子旁，令 I_j 等于 1,其中 $j=1,\cdots,10$. 那么,如果 X 表示坐在相同编号桌子旁的已婚夫妇数，有

$$X = \sum_{j=1}^{10} I_j$$

并且

$$E[X] = \sum_{j=1}^{10} E[I_j]$$

(a) 在这种情况下，为了计算 $E[I_j]$，考虑编号为 j 的妻子. 因为不包括她的、大小为 3 的$\binom{19}{3}$个组中的每一个等可能地和她坐在同一张桌子旁，因此她的丈夫和她坐在同一张桌子旁的概率为

$$\frac{\binom{1}{1}\binom{18}{2}}{\binom{19}{3}}=\frac{3}{19}$$

于是，$E[I_j]=3/19$，并且

$$E[X]=30/19$$

(b) 在这种情况下，因为在编号为 j 的妻子的桌子旁的这两个男人等可能地是 10 个男人中的任一个．他们中的任一个是她丈夫的概率是 2/10，因此

$$E[I_j]=2/10, E[X]=2$$

4. 从例 2j 知，掷骰子时，每一面都出现至少一次，所掷骰子的期望次数为 $6(1+1/2+1/3+1/4+1/5+1/6)=14.7$. 如果令 X_i 表示第 i 面出现的总次数，那么 $\sum_{i=1}^{6} X_i$ 等于所掷的总次数，我们有

$$14.7=E\left[\sum_{i=1}^{6} X_i\right]=\sum_{i=1}^{6} E[X_i]$$

由对称性，对所有的 i，$E[X_i]$是相同的，因此由上面的结果知 $E[X_1]=14.7/6=2.45$.

5. 如果翻开第 j 张红卡片时，我们赢 1 美元，令 I_j 等于 1，否则令 I_j 等于 0.（例如，如果翻开的第一张卡片是红的，I_1 将等于 1.）因此，如果 X 表示赢的总钱数，那么

$$E[X]=E\left[\sum_{j=1}^{n} I_j\right]=\sum_{j=1}^{n} E[I_j]$$

如果第 j 张红卡片在第 j 张黑卡片之前出现，那么 I_j 将等于 1，由对称性，这个事件的概率等于 1/2. 因此，$E[I_j]=1/2$，并且 $E[X]=n/2$.

6. 为了说明 $N\leqslant n-1+I$，注意到如果所有的事件都发生，那么上式两边都等于 n，然而如果并不是都发生，那么不等式变为 $N\leqslant n-1$，这种情况下很明显上式是成立的．取期望得

$$E[N]\leqslant n-1+E[I]$$

然而，如果 A_i 发生，令 I_i 等于 1，否则令 I_i 等于 0，那么

$$E[N]=E\left[\sum_{i=1}^{n} I_i\right]=\sum_{i=1}^{n} E[I_i]=\sum_{i=1}^{n} P(A_i)$$

因为 $E[I]=P(A_1,\cdots,A_n)$，所以结论成立．

7. 假设 $1,2,\cdots,n$ 依数字顺序排列，考虑选择的 k 个值．从例 2m 知，第一个值的位置等于选择的最小值，有均值 $1+\frac{n-k}{k+1}=\frac{n+1}{k+1}$.

作为一个更为正式的理由，注意到如果 $j-1$ 个最小值没有被选择，则 $X\geqslant j$. 因此，

$$P\{X\geqslant j\}=\frac{\binom{n-j+1}{k}}{\binom{n}{k}}=\frac{\binom{n-k}{j-1}}{\binom{n}{j-1}}$$

上式表明 X 和例 2m 中的随机变量有相同的分布(用符号变化，总指令数为 n，特殊指令数为 k).

8. 令 X 表示在 Sanchez 之后离开的乘客数．对 $N-1$ 个非 Sanchez 乘客随机编号，如果 Sanchez 之后第 r 个乘客离开，令 I_r 等于 1，$1\leqslant r\leqslant N-1$. 那么

$$X=\sum_{r=1}^{N-1} I_r$$

取期望得

$$E[X]=\sum_{r=1}^{N-1} P\{\text{Sanchez 之后乘客 } r \text{ 离开}\}$$

下面考虑任一个登记 k 件行李的非 Sanchez 乘客．因为登记 $k+j$ 件行李的每一个，或者这个乘客或者 Sanchez 等可能地是 $k+j$ 个中的最后一个，在 Sanchez 之后离开的那个乘客的概率为 $\frac{k}{k+j}$．因为当 $k\neq j$ 时，登记 k 件行李的非 Sanchez 乘客数为 n_k，而当 $k=j$ 时，这个数为 n_j-1，故得到

$$E[X]=\sum_k \frac{kn_k}{k+j}-\frac{1}{2}$$

9. 令边缘的任一点的附近为开始点，并延续长度 1．现在同样地选择圆周上的一个点——也就是说，它在一段长度为 x 的弧上的概率为 $\frac{x}{2\pi}$，并令 X 表示在它附近的点数．如果钉子 j 在随机点的附近令 I_j 等于 1，否则等于 0，有

$$X=\sum_{j=1}^{19} I_j$$

取期望得

$$E[X]=\sum_{j=1}^{19}P\{\text{钉子 } j \text{ 在随机点的附近}\}$$

但是因为钉子 j 将在随机点的附近，如果这个随机点在从钉子 j 开始，逆时针方向的长度为 1 的弧上，因此

$$P\{\text{钉子 } j \text{ 在随机点的附近}\}=\frac{1}{2\pi}$$

于是，

$$E[X]=\frac{19}{2\pi}>3$$

因为 $E[X]>3$ 至少有一个 X 的值必须超过 3，故结论得证．

10. 如果 $g(x)=x^{1/2}$，那么

$$g'(x)=\frac{1}{2}x^{-1/2},g''(x)=-\frac{1}{4}x^{-3/2}$$

因此 $\sqrt{x}$ 在 λ 处的泰勒级数展开式为

$$\sqrt{X}\approx\sqrt{\lambda}+\frac{1}{2}\lambda^{-1/2}(X-\lambda)-\frac{1}{8}\lambda^{-3/2}(X-\lambda)^2$$

取期望得

$$\begin{aligned}E[\sqrt{X}]&\approx\sqrt{\lambda}+\frac{1}{2}\lambda^{-1/2}E[X-\lambda]-\frac{1}{8}\lambda^{-3/2}E[(X-\lambda)^2]\\&=\sqrt{\lambda}-\frac{1}{8}\lambda^{-3/2}\lambda=\sqrt{\lambda}-\frac{1}{8}\lambda^{-1/2}\end{aligned}$$

因此，

$$\begin{aligned}\mathrm{Var}(\sqrt{X})&=E[X]-(E[\sqrt{X}])^2\approx\lambda-\left(\sqrt{\lambda}-\frac{1}{8}\lambda^{-1/2}\right)^2\\&=1/4-\frac{1}{64\lambda}\approx1/4\end{aligned}$$

11. 给桌子编号，桌子 1，2，3 为有 4 个座位的桌子，桌子 4，5，6，7 为有 2 个座位的桌子．另外，给妻子编号，如果编号为 i 的妻子与她丈夫坐在编号为 j 的桌子旁，令 X_{ij} 等于 1．注意到

$$E[X_{i,j}]=\frac{\binom{2}{2}\binom{18}{2}}{\binom{20}{4}}=\frac{3}{95},\ j=1,2,3$$

和

$$E[X_{i,j}]=\frac{1}{\binom{20}{2}}=\frac{1}{190},\ j=4,5,6,7$$

令 X 表示坐在同一张桌子旁的夫妇数，于是我们有

$$E[X]=E\left[\sum_{i=1}^{5}\sum_{j=1}^{7}X_{i,j}\right]$$
$$=\sum_{i=1}^{5}\sum_{j=1}^{3}E[X_{i,j}]+\sum_{i=1}^{5}\sum_{j=4}^{7}E[X_{i,j}]$$
$$=15(3/95)+20(1/190)=11/19$$

12. 如果第 i 号人没有作为招聘者，令 $X_i=1$，否则令 $X_i=0$，则

$$E[X_i]=P\{i\text{ 没有招聘 } i+1,i+2,\cdots,n\}=\frac{i-1}{i}\frac{i}{i+1}\cdots\frac{n-2}{n-1}=\frac{i-1}{n-1}$$

因此，

$$E\left[\sum_{i=1}^{n}X_i\right]=\sum_{i=1}^{n}\frac{i-1}{n-1}=\frac{n}{2}$$

从上式可以得到

$$\mathrm{Var}(X_i)=\frac{i-1}{n-1}\left(1-\frac{i-1}{n-1}\right)=\frac{(i-1)(n-i)}{(n-1)^2}$$

对 $i<j$，

$$E[X_iX_j]=\frac{i-1}{i}\cdots\frac{j-2}{j-1}\frac{j-2}{j}\frac{j-1}{j+1}\cdots\frac{n-3}{n-1}=\frac{(i-1)(j-2)}{(n-2)(n-1)}$$

并且

$$\mathrm{Cov}(X_i,X_j)=\frac{(i-1)(j-2)}{(n-2)(n-1)}-\frac{i-1}{n-1}\frac{j-1}{n-1}=\frac{(i-1)(j-n)}{(n-2)(n-1)^2}$$

因此，

$$\mathrm{Var}\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}\mathrm{Var}(X_i)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathrm{Cov}(X_i,X_j)$$
$$=\sum_{i=1}^{n}\frac{(i-1)(n-i)}{(n-1)^2}+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{(i-1)(j-n)}{(n-2)(n-1)^2}$$
$$=\frac{1}{(n-1)^2}\sum_{i=1}^{n}(i-1)(n-i)-\frac{1}{(n-2)(n-1)^2}\sum_{i=1}^{n-1}(i-1)(n-i)(n-i-1)$$

13. 如果第 i 组包含所有三种类型的球员中的一个，令 X_i 等于 1，那么

$$E[X_i]=\frac{\binom{2}{1}\binom{3}{1}\binom{4}{1}}{\binom{9}{3}}=\frac{2}{7}$$

因此，对(a)得到

$$E\left[\sum_{i=1}^{3}X_i\right]=6/7$$

从上面的式子得到

$$\mathrm{Var}(X_i)=(2/7)(1-2/7)=10/49$$

并且，对 $i\neq j$，

$$E[X_iX_j]=P\{X_i=1,X_j=1\}$$
$$=P\{X_i=1\}P\{X_j=1|X_i=1\}$$
$$=\frac{\binom{2}{1}\binom{3}{1}\binom{4}{1}\binom{1}{1}\binom{2}{1}\binom{3}{1}}{\binom{9}{3}\quad\binom{6}{3}}$$
$$=6/70$$

因此，对于(b)得到

$$\begin{aligned}\mathrm{Var}\Big(\sum_{i=1}^{3}X_i\Big) &= \sum_{i=1}^{3}\mathrm{Var}(X_i)+2\sum\sum_{j>1}\mathrm{Cov}(X_i,X_j)\\ &=30/49+2\binom{3}{2}\Big(\frac{6}{70}-\frac{4}{49}\Big)\\ &=\frac{312}{490}\end{aligned}$$

14. 如果第 i 张牌是 A，令 X_i 等于 1，否则令 X_i 等于 0；如果第 j 张是黑桃，令 Y_j 等于 1，否则令 Y_j 等于 0，$i,j=1,\cdots,13$. 于是得到

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{Cov}\Big(\sum_{i=1}^{n}X_i,\sum_{j=1}^{n}Y_j\Big)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i,Y_j)\end{aligned}$$

然而，很明显，X_i 和 Y_j 是独立的，因为一张指定的牌的花色与它是否是 A 无关，从而不影响另一张牌是 A 的概率．更正式地，令 $A_{i,s},A_{i,h},A_{i,d},A_{i,c}$ 分别表示第 i 张牌是黑桃，红桃，方块，梅花的事件，那么

$$P\{Y_j=1\}=\frac{1}{4}(P\{Y_j=1|A_{i,s}\}+P\{Y_j=1|A_{i,h}\}+P\{Y_j=1|A_{i,d}\}+P\{Y_j=1|A_{i,c}\})$$

然而，由对称性得

$$P\{Y_j=1|A_{i,s}\}=P\{Y_j=1|A_{i,h}\}=P\{Y_j=1|A_{i,d}\}=P\{Y_j=1|A_{i,c}\}$$

因此，

$$P\{Y_j=1\}=P\{Y_j=1|A_{i,s}\}$$

上式暗含着

$$P\{Y_j=1\}=P\{Y_j=1|A_{i,s}^c\}$$

可以看到 X_i 和 Y_j 是独立的．因此，$\mathrm{Cov}(X_i,Y_j)=0$，且有 $\mathrm{Cov}(X,Y)=0$. 随机变量 X 和 Y 虽然是不相关的，但并不独立．这一点由下面的特例可得到：

$$P\{Y=13|X=4\}=0\neq P\{Y=13\}$$

15. (a) 没有任何信息时，赢得钱数的期望值是 0.

(b) 如果 $p>1/2$ 预测出现正面，否则，预测出现反面．

(c) 以赢得钱数 V 为条件，得到的期望值为

$$\begin{aligned}E[\text{获得}] &= \int_0^1 E[\text{获得}\mid V=p]\mathrm{d}p\\ &= \int_0^{1/2}[1(1-p)-1(p)]\mathrm{d}p+\int_{1/2}^{1}[1(p)-1(1-p)]\mathrm{d}p = 1/2\end{aligned}$$

16. 考虑到被选择的名字出现在表中的 $n(X)$ 个不同位置，因为这些位置中的每一个等可能地是被选择的这一个，得

$$E[I|n(X)]=P\{I=1|n(X)\}=1/n(X)$$

因此，

$$E[I]=E[1/n(X)]$$

并且，$E[mI]=E[m/n(X)]=d$.

17. 如果放第 i 条指令时一个冲突发生，令 X_i 等于 1，否则令 X_i 等于 0，冲突总数 X 可以表示为

$$X=\sum_{i=1}^{m}X_i$$

因此，

$$E[X]=\sum_{i=1}^{m}E[X_i]$$

为了确定 $E[X_i]$，以被放的单元为条件．

$$\begin{aligned}E[X_i] &= \sum_j E[X_i \mid \text{放在单元 } j]p_j \\ &= \sum_j P\{i \text{ 导致冲突} \mid \text{放在单元 } j\}p_j \\ &= \sum_j [1-(1-p_j)^{i-1}]p_j \\ &= 1-\sum_j (1-p_j)^{i-1}p_j\end{aligned}$$

倒数第二个等式用到了：以第 i 条指令放在单元 j 为条件，如果上述 $i-1$ 条指令中的任一条被放在单元 j，那么第 i 条指令将导致一个冲突．因此，

$$E[X] = m - \sum_{i=1}^{m}\sum_{j=1}^{n}(1-p_j)^{i-1}p_j$$

互换求和的顺序，得

$$E[X] = m-n+\sum_{j=1}^{n}(1-p_j)^m$$

观察上式表明，上式可以通过下式两边求期望更容易地得到：

$$\text{非空单元数} = m - X$$

通过对每一个单元定义一个示性变量可以得到非空单元的期望数，如果单元非空令其等于 1，否则令其等于 0，然后取这些示性变量之和的期望值．

18. 令 L 表示第一个游程的长度．以第一个值为条件得

$$E[L]=E[L\mid\text{第一个值是 }1]\frac{n}{n+m}+E[L\mid\text{第一个值是 }0]\frac{m}{n+m}$$

如果第一个值是 1，那么当考虑剩下的 $n+m-1$ 个值(其中 $n-1$ 个 1 和 m 个 0)时，游程的长度为第一个 0 所在的位置．(例如，如果剩下的 $n+m-1$ 个值中开始值是 0，那么 $L=1$．)对于第一个值是 0 的情况有相似的结论，利用上述结论，并运用例 2m 的结论可以得到

$$E[L]=\frac{n+m}{m+1}\frac{n}{n+m}+\frac{n+m}{n+1}\frac{m}{n+m}=\frac{n}{m+1}+\frac{m}{n+1}$$

19. 令 X 表示两个箱子都空时抛硬币的次数，令 Y 表示在前 $n+m$ 次抛硬币中正面出现的次数，那么

$$E[X]=\sum_{i=0}^{n+m}E[X\mid Y=i]P\{Y=i\}=\sum_{i=0}^{n+m}E[X\mid Y=i]\binom{n+m}{i}p^i(1-p)^{n+m-i}$$

如果在前 $n+m$ 次抛硬币中出现的正面数为 i，$i\leqslant n$，那么还需要抛硬币的次数是为了得到另外的 $n-i$ 次正面．同样，如果在前 $n+m$ 次抛硬币中出现的正面数为 i，$i>n$，那么已经有 $n+m-i<m$ 次反面，另外还需要抛硬币的次数是为了得到 $i-n$ 次正面．由于对于 j 次特定类型的结果需要的抛硬币次数是负二项随机变量，均值为 j 除以那种结果的概率，因此得到

$$E[X]=\sum_{i=0}^{n}\frac{n-i}{p}\binom{n+m}{i}p^i(1-p)^{n+m-i}+\sum_{i=n+1}^{n+m}\frac{i-n}{1-p}\binom{n+m}{i}p^i(1-p)^{n+m-i}$$

20. 在给出的提示的等式两边取期望得

$$\begin{aligned}E[X^n] &= E\left[n\int_0^\infty x^{n-1}I_X(x)\mathrm{d}x\right]=n\int_0^\infty E[x^{n-1}I_X(x)]\mathrm{d}x \\ &= n\int_0^\infty x^{n-1}E[I_X(x)]\mathrm{d}x = n\int_0^\infty x^{n-1}\overline{F}(x)\mathrm{d}x\end{aligned}$$

在积分号内取期望是合理的，因为所有的随机变量 $I_X(x)(0<x<\infty)$都是非负的．

第 8 章

1. 令 X 表示下周的销售量，注意 X 是整数．由马尔可夫不等式得到

(a) $P\{X>18\}=P\{X\geqslant 19\}\leqslant\frac{E[X]}{19}=16/19$

(b) $P\{X>25\}=P\{X\geqslant 26\}\leqslant\frac{E[X]}{26}=16/26$

2. (a) $P\{10\leqslant X\leqslant 22\}=P\{|X-16|\leqslant 6\}=P\{|X-\mu|\leqslant 6\}$
$=1-P\{|X-\mu|>6\}\geqslant 1-9/36=3/4$

(b) $P\{X\geqslant 19\}=P\{X-16\geqslant 3\}\leqslant\frac{9}{9+9}=1/2$

在(a)中用到了切比雪夫不等式，在(b)中是单边的情况(参看命题 5.1).

3. 首先注意 $E[X-Y]=0$，并且

$$\mathrm{Var}(X-Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)-2\mathrm{Cov}(X,Y)=28$$

在(a)中用切比雪夫不等式，在(b)和(c)中用单边的情况，得到下面的结论：

(a) $P\{|X-Y|>15\}\leqslant 28/225$

(b) $P\{X-Y>15\}\leqslant\frac{28}{28+225}=28/253$

(c) $P\{Y-X>15\}\leqslant\frac{28}{28+225}=28/253$

4. 如果 X 表示工厂 A 的产量，Y 表示工厂 B 的产量，那么

$$E[Y-X]=-2,\ \mathrm{Var}(Y-X)=36+9=45$$

$$P\{Y-X>0\}=P\{Y-X\geqslant 1\}=P\{Y-X+2\geqslant 3\}\leqslant\frac{45}{45+9}=45/54$$

5. 首先注意

$$E[X_i]=\int_0^1 2x^2\,\mathrm{d}x=2/3$$

运用强大数定律得到

$$r=\lim_{n\to\infty}\frac{n}{S_n}=\lim_{n\to\infty}\frac{1}{S_n/n}=\frac{1}{\lim\limits_{n\to\infty}S_n/n}=1/(2/3)=3/2$$

6. 因为 $E[X_i]=2/3$，并且

$$E[X_i^2]=\int_0^1 2x^3\,\mathrm{d}x=1/2$$

得到 $\mathrm{Var}(X_i)=1/2-(2/3)^2=1/18$. 于是，如果手头有 n 个元件，则

$$\begin{aligned}P\{S_n\geqslant 35\}&=P\{S_n\geqslant 34.5\}\quad\text{(连续性修正)}\\&=P\left\{\frac{S_n-2n/3}{\sqrt{n/18}}\leqslant\frac{34.5-2n/3}{\sqrt{n/18}}\right\}\\&\approx P\left\{Z\geqslant\frac{34.5-2n/3}{\sqrt{n/18}}\right\}\end{aligned}$$

其中 Z 是一个标准正态随机变量. 由于

$$P\{Z>-1.284\}=P\{Z<1.284\}\approx 0.90$$

可以看到应选择 n 使得

$$(34.5-2n/3)\approx -1.284\sqrt{n/18}$$

计算可得 $n=55$.

7. 如果 X 表示维修机器的时间，则

$$E[X]=0.2+0.3=0.5$$

并且运用指数分布随机变量的方差等于均值的平方，得到

$$\mathrm{Var}(X)=(0.2)^2+(0.3)^2=0.13$$

所以，用 X_i 表示维修 i 的时间，$i=1,\cdots,20$，并且 Z 表示一个标准正态随机变量，

$$P\{X_1+\cdots+X_{20}<8\}=P\left\{\frac{X_1+\cdots+X_{20}-10}{\sqrt{2.6}}<\frac{8-10}{\sqrt{2.6}}\right\}$$

$$\approx P\{Z<-1.240\ 35\}\approx 0.107\ 4$$

8. 首先注意，如果 X 表示一次赌博中赌徒得到的奖金，那么

$$E[X]=-0.7-0.4+1=-0.1,E[X^2]=0.7+0.8+10=11.5\rightarrow \text{Var}(X)=11.49$$

因此，由于 Z 服从标准正态分布，

$$P\{X_1+\cdots+X_{100}\leqslant -0.5\}=P\left\{\frac{X_1+\cdots+X_{100}+10}{\sqrt{1\ 149}}\leqslant\frac{-0.5+10}{\sqrt{1\ 149}}\right\}$$
$$\approx P\{Z\leqslant 0.280\ 3\}\approx 0.610\ 4$$

9. 用自测题 7 中的记号，有

$$P\{X_1+\cdots+X_{20}<t\}=P\left\{\frac{X_1+\cdots+X_{20}-10}{\sqrt{2.6}}<\frac{t-10}{\sqrt{2.6}}\right\}\approx P\left\{Z<\frac{t-10}{\sqrt{2.6}}\right\}$$

由于 $P\{Z<1.645\}\approx 0.95$，因此 t 应该满足

$$\frac{t-10}{\sqrt{2.6}}\approx 1.645$$

得到 $t\approx 12.65$.

10. 如果这个声明是正确的，那么由中心极限定理得到：平均的尼古丁含量（记为 X）近似服从均值为 2.2、标准差为 0.03 的正态分布．因此它大于 3.1 的概率为

$$P\{X>3.1\}=P\left\{\frac{X-2.2}{\sqrt{0.03}}>\frac{3.1-2.2}{\sqrt{0.03}}\right\}\approx P\{Z>5.196\}\approx 0$$

此处 Z 是一个标准正态随机变量．

11. 令 N 表示志愿医生数．以事件 $N=i$ 为条件，病人数被认为是 i 个均值为 30 的独立泊松随机变量之和．因为独立泊松随机变量之和仍然是泊松随机变量，因此给定 $N=i$ 的条件下，X 的条件分布为均值为 $30i$ 的泊松分布．于是，

$$E[X|N]=30N\quad \text{Var}(X|N)=30N$$

结果，

$$E[X]=E[E[X|N]]=30E[N]=90$$

并且，由条件方差公式，

$$\text{Var}(X)=E[\text{Var}(X|N)]+\text{Var}(E[X|N])=30E[N]+(30)^2\text{Var}(N)$$

由

$$\text{Var}(N)=\frac{1}{3}(2^2+3^2+4^2)-9=2/3$$

得到 $\text{Var}(X)=690$.

为了近似 $P\{X>65\}$，没有理由假设 X 近似服从均值为 90、方差为 690 的正态分布．然而，我们确实知道

$$P\{X>65\}=\sum_{i=2}^{4}P\{X>65\mid N=i\}P\{N=i\}=\frac{1}{3}\sum_{i=2}^{4}\overline{P}_i(65)$$

此处 $\overline{P}_i(65)$ 是均值为 $30i$ 的泊松随机变量大于 65 的概率．也就是说，

$$\overline{P}_i(65)=1-\sum_{j=0}^{65}e^{-30i}(30i)^j/j!$$

因为均值为 $30i$ 的泊松随机变量与 $30i$ 个均值为 1 的泊松随机变量之和有相同的分布，因此由中心极限定理知，它的分布近似服从均值和方差都等于 $30i$ 的正态分布．从而，用 X_i 表示均值为 $30i$ 的泊松随机变量，Z 表示标准正态随机变量，可以如下近似 $\overline{P}_i(65)$：

$$\overline{P}_i(65)=P\{X>65\}=P\{X\geqslant 65.5\}$$
$$=P\left\{\frac{X-30i}{\sqrt{30i}}\geqslant\frac{65.5-30i}{\sqrt{30i}}\right\}\approx P\left\{Z\geqslant\frac{65.5-30i}{\sqrt{30i}}\right\}$$

因此，

$$\overline{P}_2(65)\approx P\{Z\geqslant 0.710\,0\}\approx 0.238\,9$$
$$\overline{P}_3(65)\approx P\{Z\geqslant -2.583\}\approx 0.995\,1$$
$$\overline{P}_4(65)\approx P\{Z\geqslant -4.975\}\approx 1$$

得到结果为

$$P\{X>65\}\approx 0.744\,7$$

如果已经错误地假设 X 近似是正态的，将得到近似答案为 0.824 4.（正确答案为0.744 0.）

第 9 章

1. 由公理(ⅲ)得，发生在 8 点和 10 点间的事件数与到 2 点发生的事件数有相同的分布，都是均值为 6 的泊松随机变量．因此，可得(a)和(b)的答案．

(a) $P\{N(10)-N(8)=0\}=e^{-6}$

(b) $E[N(10)-N(8)]=6$

(c) 由公理(ⅱ)和(ⅲ)得，发生在任一时间点以前的事件过程都是强度为 λ 的泊松过程．因此，下午 2 点之后第 5 个事件的期望时间是 $2+E[S_5]=2+5/3$. 也就是说，这个事件的期望时间是下午 3:40.

2. (a) $P\{N(1/3)=2|N(1)=2\}$

$$=\frac{P\{N(1/3)=2,N(1)=2\}}{P\{N(1)=2\}}$$
$$=\frac{P\{N(1/3)=2,N(1)-N(1/3)=0\}}{P\{N(1)=2\}}$$
$$=\frac{P\{N(1/3)=2\}P\{N(1)-N(1/3)=0\}}{P\{N(1)=2\}}\quad(\text{由公理(ⅱ)})$$
$$=\frac{P\{N(1/3)=2\}P\{N(2/3)=0\}}{P\{N(1)=2\}}\quad(\text{由公理(ⅲ)})$$
$$=\frac{e^{-\lambda/3}(\lambda/3)^2/2!\ e^{-2\lambda/3}}{e^{-\lambda}\lambda^2/2!}=1/9$$

(b) $P\{N(1/2)\geqslant 1|N(1)=2\}=1-P\{N(1/2)=0|N(1)=2\}$

$$=1-\frac{P\{N(1/2)=0,N(1)=2\}}{P\{N(1)=2\}}$$
$$=1-\frac{P\{N(1/2)=0,N(1)-N(1/2)=2\}}{P\{N(1)=2\}}$$
$$=1-\frac{P\{N(1/2)=0\}P\{N(1)-N(1/2)=2\}}{P\{N(1)=2\}}$$
$$=1-\frac{P\{N(1/2)=0\}P\{N(1/2)=2\}}{P\{N(1)=2\}}$$
$$=1-\frac{e^{-\lambda/2}e^{-\lambda/2}(\lambda/2)^2/2!}{e^{-\lambda}\lambda^2/2!}$$
$$=1-1/4=3/4$$

3. 在路上固定一个点，如果经过的第 n 辆车是汽车令 X_n 等于 0，如果是卡车令 X_n 等于 1，$n\geqslant 1$. 现在假设序列 $X_n(n\geqslant 1)$是一个马尔可夫链，其转移概率为

$$P_{0,0}=5/6,P_{0,1}=1/6,P_{1,0}=4/5,P_{1,1}=1/5$$

因此，这个链的长远分布满足

$$\pi_0=\pi_0(5/6)+\pi_1(4/5)$$
$$\pi_1=\pi_0(1/6)+\pi_1(1/5)$$
$$\pi_0+\pi_1=1$$

解上述方程得

$$\pi_0=24/29\quad \pi_1=5/29$$

因此，在路上有 24/29≈83%是汽车．

4. 连续的天气分类构成一个马尔可夫链．如果阴天状态为 0，晴天状态为 1，多云状态为 2，那么转移概率矩阵如下：

$$\boldsymbol{P}=\begin{matrix}0 & 1/2 & 1/2\\ 1/3 & 1/3 & 1/3\\ 1/3 & 1/3 & 1/3\end{matrix}$$

这个链的长远分布满足

$$\pi_0=\pi_1(1/3)+\pi_2(1/3)$$
$$\pi_1=\pi_0(1/2)+\pi_1(1/3)+\pi_2(1/3)$$
$$\pi_2=\pi_0(1/2)+\pi_1(1/3)+\pi_2(1/3)$$
$$1=\pi_0+\pi_1+\pi_2$$

上述方程的解为

$$\pi_0=1/4,\pi_1=3/8,\pi_2=3/8$$

因此，晴天占 3/8，阴天占 1/4.

5. (a) 直接计算得到

$$H(X)/H(Y)\approx 1.06$$

(b) 两个随机变量以相同的概率 0.35 和 0.05 取两个值．不同的是，如果它们不取那些值中的任一个，那么 X 而不是 Y 等可能地取它其他的三个值中的任一个．因此，由习题 13 将得到(a)中的结果．

第 10 章

1. (a) $1=C\int_0^1 e^x dx \Rightarrow C=1/(e-1)$.

(b) $F(x)=C\int_0^x e^y dy=\frac{e^x-1}{e-1}, 0\leqslant x\leqslant 1$.

因此，如果令 $X=F^{-1}(U)$，那么

$$U=\frac{e^X-1}{e-1}$$

或者

$$X=\log(U(e-1)+1)$$

于是，可以通过产生一个随机数 U，并令 $X=\log(U(e-1)+1)$来模拟随机变量 X.

2. 利用接受-拒绝法，其中 $g(x)=1(0<x<1)$．计算表明 $f(x)/g(x)$在某个 $x(0<x<1)$达到最大值，x 满足

$$2x-6x^2+4x^3=0$$

或者，等价地，

$$4x^2-6x+2=(4x-2)(x-1)=0$$

因此，当 $x=1/2$ 时达到最大值，并且

$$C=\max f(x)/g(x)=30(1/4-2/8+1/16)=15/8$$

因此，算法如下：

步骤 1. 产生一个随机数 U_1.

步骤 2. 产生一个随机数 U_2.

步骤 3. 如果 $U_2\leqslant 16(U_1^2-2U_1^3+U_1^4)$，令 $X=U_1$；否则转到步骤 1.

3. 首先检查较高的概率值，下面的算法是有效的：

步骤 1. 产生一个随机数 U.

步骤 2. 如果 $U\leqslant 0.35$，令 $X=3$ 并停止．

步骤 3. 如果 $U\leqslant 0.65$，令 $X=4$ 并停止．

步骤 4. 如果 $U\leqslant 0.85$，令 $X=2$ 并停止．

步骤 5. $X=1$.

4. $2\mu - X$

5. (a) 产生 $2n$ 个均值为 1 的独立指数随机变量 $X_i, Y_i, i=1,\cdots,n$，然后用估计值 $\sum_{i=1}^{n} e^{X_i Y_i}/n$.

(b) 用 XY 作为一个控制变量获得形式为下式的一个估计值：

$$\sum_{i=1}^{n}(e^{X_i Y_i} + cX_i Y_i)/n$$

另一种可能是用 $XY + X^2Y^2/2$ 作为控制变量，获得形式为下式的一个估计值：

$$\sum_{i=1}^{n}(e^{X_i Y_i} + c[X_i Y_i + X_i^2 Y_i^2/2 - 1/2])/n$$

上式基于下列事实：e^{xy} 的麦克劳林级数展开式的前三项为

$$1 + xy + (x^2 y^2)/2$$

索　引

索引中的页码为英文原书页码，与书中页边标注的页码一致.

N

O

P

Q

R

S

T

U

V

W

Z